IDA

INSTITUTE FOR DEFENSE ANALYSES

Pathways to Cooperation between the Intelligence Community and the Social and Behavioral Science Communities

Steven E. Koonin
Sallie A. Keller
Stephanie S. Shipp
Ted W. Allen
Gina K. Walejko

March 2013

Approved for public release; distribution is unlimited.

IDA Paper P-5000

Log: H 13-000361

INSTITUTE FOR DEFENSE ANALYSES
4850 Mark Center Drive
Alexandria, Virginia 22311-1882

About This Publication

This work was conducted by the Institute for Defense Analyses (IDA) under contract DASW01-04-C-0003, Task ET-20-3393, "ODNI Scientific and Technical (S&T) Capabilities at Federal Labs and S&T for the Intelligence Community," for the Office of the Director of National Intelligence (ODNI). The views, opinions, and findings should not be construed as representing the official position of either ODNI or the Department of Defense.

Acknowledgments

The technical reviewers for this work were Robert E. Roberts, Chief Scientist of IDA and founder and director of the Defense Sciences Study Group; Sue K. Numrich of IDA; IDA consultants Robert H. Latiff and Kenan P. Jarboe; and David Roessner of SRI International.

INSTITUTE FOR DEFENSE ANALYSES

IDA Paper P-5000

Pathways to Cooperation between the Intelligence Community and the Social and Behavioral Science Communities

Steven E. Koonin
Sallie A. Keller
Stephanie S. Shipp
Ted W. Allen
Gina K. Walejko

Executive Summary

Introduction

This report proposes a community of practice between the U.S. Intelligence Community (IC) and social and behavioral scientists for the purpose of addressing issues of importance to the IC, researching topics relevant to social and behavioral scientists, and, where appropriate, focusing attention on the intelligent management and use of big data. The central goal of this report is to highlight potential options for developing cooperative programs by investigating the history, issues, and implications of some past interactions between the two communities. Crucial to developing those programs is finding junctures where the interests of the communities overlap sufficiently to create significant benefits for both.

Two such junctures are the “big data” created by the deployment of technology-based infrastructures and the trends in implementing automated systems for smart cities. For these cooperative endeavors to succeed, it will be necessary to address fundamental hurdles. One such hurdle is the issue of privacy. As more and more data are collected on such large scales for use by the government, the private sector, and others, establishing guidelines for handling big data will be a necessary step. Some solutions are already in place and others will be developed. At the same time, some studies have found that when the economic benefits of convenience exceed the costs of privacy, many are willing to give up some personal privacy.

The premise for this paper is that the IC and social science community could benefit by working together on these topics of mutual interest (for example, the use of big data and smart cities) in the context of developing new statistical methods and tools and analyses of topics of mutual interest using unclassified data initially and as trust is built, classified data. The goal is that these collaborations would benefit both communities without jeopardizing the identity and culture of each.

On May 11, 2012, the Institute for Defense Analyses hosted the Big Data for Social Sciences workshop as a first step to discuss the fundamentals of applying big data to social science issues and methods for collaborating on big data research.

Building a Community of Practice

An unprecedented forum for collaboration, a group of civilian scientists, and continued funding are the necessary components for building a community of practice between the IC and social and behavioral scientists. The report explores these components by

- examining how to establish a precedent for IC collaboration with a group of civilian scientists, using as an example the Measurements of Earth Data for Environmental Analysis (MEDEA) whereby the IC released data and images to social scientists for use in environmental analyses and gained trust of the civilian science community;
- discussing how to build a community of civilian scientists and introducing them to aspects of the Federal Government, using as an example the Defense Science Study Group (DSSG), which the Defense Advanced Research Projects Agency (DARPA) developed to bring together the national security community and emerging leaders in academia in science and technology disciplines;
- proposing the establishment of continued funding for civilian scientists, examining as possible models the Minerva Research Initiative, a Department of Defense (DOD) program to fund social and behavioral science research in areas strategically important to U.S. national security, the Intelligence Advanced Research Projects Activity (IARPA), or the National Institute of Standards and Technology (NIST) Advanced Technology Program; and
- creating an advisory board of prominent social and behavioral scientists, including computational statisticians.

The report discusses how the example programs might or might not be adapted to address IC issues. Ideas for new programs take into account the lessons learned from earlier programs. These programs could provide a test bed for developing new tools, methods, and analytical findings for exploring the use of big data to study smart cities, which is a potential research topic of interest to both the IC and civilian scientists.

Establish a Precedent

Perhaps one of the IC's better known collaborations with the open scientific community was the establishment of MEDEA in 1993. Under MEDEA about 60 environmental scientists were given access to selected classified documents and hundreds of thousands of satellite images to use in studying environmental problems. The MEDEA collaboration was successful for the scientific community in particular and for society more broadly. For example, MEDEA resulted in not only advances in climate science, but also in archaeological discoveries that advanced researchers' understanding of ancient civilizations. As a result, the program helped build trust between civilian scientists and the IC. In addition, MEDEA brought scientists into the IC where they first worked with IC data. This led to the declassification of data and release to the broader research community. This model could still work if there are data that the IC can share with researchers in either a classified or unclassified setting.

The specific data the IC may or may not be able to share is undefined right now, which parallels the uncertainty surrounding the start of MEDEA. The goal is to establish a precedent so that advances in research on big data will have spillovers to multiple communities. The aim is to create and build methods and tools to use new sources of data that are big, messy, unknown, and potentially useful. This would involve risk in that there are likely to be more failures than successes, at least initially. However, failures can be instructive, and if the research is truly high risk, the payback from just one or two successes is likely to exceed the total cost of the program.

Build a Community

A successful collaboration between the DOD and academia is found in the 25-year-old DSSG program. This program identifies emerging science and technology leaders with the goals to educate them about the national security community and its challenges, foster in them a long-term interest in national security, and facilitate their involvement on behalf of the national security community.

The January 2012 *Report of the Defense Science Board Task Force on Basic Research* highlights the need for attracting the participation of the nation's top scientists and engineers in national security challenges. It singles out the DSSG as a unique program that meets that challenge and recommends expansion of the program to also include social and behavioral scientists.

Create a Funding Source

One of the defense community's more recent ventures into the social and behavioral sciences has been the Minerva Research Initiative. The DOD is providing $75 million over 5 years to fund research in areas strategically important to U.S. national security. The initiative has drawn criticism from academic communities for blurring the lines between the military and science and using scientists to further national security goals.

The IARPA was created in 2008, and according to its website, "invests in high-risk, high-payoff research programs that have the potential to provide the United States with an overwhelming intelligence advantage over future adversaries."[1] While the IC's 17 intelligence agencies and organizations focus on the day-to-day concerns within their own particular sphere of operation, IARPA tackles research that spans across or fills gaps between intelligence agencies. The primary expectation is for IARPA to create

---

[1] IARPA – Our Organization, http://www.iarpa.gov/organization.html.

transformational solutions that can be transitioned to the IC, but one of the major attractions for IARPA is the expected spillover solutions that benefit the civilian population. A similar program, the NIST Advanced Technology Program was less prescriptive than the current IARPA program. Both models could be examined to develop a program that would benefit the IC and other communities more broadly.

Create Visibility and Transparency in Research

One approach to setting up this engagement is to bring together an advisory group composed of IC experts and prominent statisticians, social and behavioral scientists, and physical scientists to create a plan for the communities to work together. The Defense Science Board and the Air Force Scientific Advisory Board, for example, invite well-known leaders to actively participate in their proceedings and provide guidance and research. These boards create credibility, visibility, and excitement for specific areas of research.

Potential Activities to Build a Community of Practice

Big data offers opportunities for productive collaboration between the IC and social and behavioral scientists. An optimal blend of some components of MEDEA, DSSG, IARPA, and other models would allow researchers access to IC data for research that may identify potentially transformative events and advance the IC's mission to be looking outward toward what other countries are doing and forward toward what will happen over the next 20 years. Lessons from these and other previous collaborations suggest approaches for new programs that could provide benefits to the public, the social and behavioral sciences, and the IC.

One avenue for the IC's collaboration with social and behavioral scientists would be in the application of existing technologies and methodologies the IC typically uses to leverage the development of a variety of "smart" or "intelligent" cities in selected U.S. locations and to better understand cities overseas. Smart cities have sparked interest around the world as a way to improve urban efficiency, resilience, and sustainability. To accomplish these goals, the smart city enterprise must have the technological ability through sensing and communications systems to (1) collect observational data from the infrastructure, environment, and people; (2) get the data to the right place at the right time; and (3) integrate, analyze, and act upon data collected from diverse sources. Social scientists would develop new computational methods and address other big data issues using the smart city as a platform. The IC community would benefit by having access to these new methods and approaches, as well as important findings about cities. In the process, the IC could also gain invaluable experience in understanding how to observe cities of the future. Questions remain about what types of interactions might help overcome barriers to studying foreign cities, especially those that may be less likely to collaborate with the United States, such as Beijing and Shanghai, Iran, or South Korea.

Summary and Recommendations

This report proposes that the IC increase its interactions and collaborations with social and behavioral scientists to share knowledge and perspectives on issues of mutual interest that would benefit both the intelligence and scientific communities. By learning from existing programs, the IC might pursue the following approaches to building a community of big data users, advisors, and researchers:

- ***Establish a precedent***. Create a MEDEA counterpart for the social and behavioral sciences to set a precedent for working together on topics of mutual interest, such as understanding cities today and as they evolve in the future.
- ***Build a community***. Sponsor a social and behavioral sciences study group modeled on the DSSG. Such a program would build a community of social and behavioral scientists that are knowledgeable about the Federal Government and available to help the IC. They would also participate in the building of methods, tools, data sources, and analyses of big data.
- ***Create a funding source***. Expand IARPA funding programs in the social and behavioral sciences, with the goal to develop new methods and approaches for using big data sources to address these and social and behavioral science issues. This program would provide funding to the academic communities on specific topics, while avoiding the pitfalls of the Minerva Research Initiative.
- ***Create visibility and transparency in research***. One approach to setting up this engagement is to bring together an advisory group composed of IC experts and prominent social and behavioral scientists, statisticians, and physical scientists to create a plan for the communities to work together.

In creating these programs, the IC can learn from the successes of MEDEA, DSSG, and IARPA and the concerns related to the Minerva Initiative. Regardless of the nature of the engagement, an agreement must be worked out so that both sides benefit.

Contents

1. Introduction

A. Study Motivation and Purpose

A community of practice between the U.S. Intelligence Community (IC) and social and behavioral scientists is proposed for the purpose of addressing issues of importance to the IC, researching topics relevant to social and behavioral scientists, and, where appropriate, focusing attention on the intelligent management and use of big data. The analyses of these “big data” derived from digital technology will require new approaches. Collaborations with social and behavioral scientists will allow the IC to build analysis capacity in new directions and with new communities. This will allow the IC to build future capability that can be used to address unknown problems and integrate information in new, yet-to-be-determined ways.

This report explores ways to elicit and integrate the cooperation of civilian social and behavioral scientists into IC and Department of Defense (DOD) security efforts. The document investigates a variety of programs and possible pathways for creating cooperative endeavors between the IC and the social and behavioral science communities.

The central motivation for this research was the common interest that these communities share in understanding and predicting behavior through the use of large datasets generated by a variety of private sources, such as those created through banking transactions, cell phone usage, ground positioning systems, and other sources. Access to these large datasets—frequently referred to as “big data”—offers social and behavioral scientists the potential opportunity to take on many of today’s most challenging societal issues. But access also requires the refinement of innovative methods and theories in addition to the integration and interpretation of such large-scale databases. All three communities could mutually benefit by collaborating in this arena.

Institute for Defense Analyses (IDA) researchers investigated ways in which the social and behavioral science community might interact with the IC by examining past and current examples of IC and DOD interactions with civilian science communities to answer the following questions:

1. How could the IC and DOD and the social and behavioral science communities collaborate? Specific examples are the Defense Science Study Group (DSSG), the Minerva Research Initiative, and the Intelligence Advanced Research Projects Activity (IARPA).

2. How could the IC set a precedent for interacting with the social and behavioral science communities using big data? One model is Measurements of Earth Data for Environmental Analysis (MEDEA).
3. How can the IC and social and behavioral scientists build a community focused on topics of mutual interest? One model is DSSG.
4. How can the IC fund civilian social and behavioral scientists? One cautionary model is the Minerva Research Initiative, and another potential model is IARPA or perhaps the former NIST Advanced Technology Program.
5. What research topics might benefit both the IC and the social and behavioral research communities? Both may benefit from access to and analysis of big data associated with "smart cities," a focus of a workshop held as part of this project.

B. Defining Big Data

Increasingly ubiquitous communication networks and technologies for sensing the physical environment, coupled with the ability to capture data from these systems and technologies in nearly real time, have produced a massive and growing collection of datasets that document human behavior and measure the conditions and events in the physical environment. Referred to as "big data," this collection of information holds the potential for intelligent management of human interaction with the physical, social, and economic environment. It also presents tremendous challenges in data management, analysis, model building, and theory generation.

Big data encompasses (1) data collected in traditional ways through purpose-guided data-gathering efforts (such as surveys, field observations, and specific-purpose physical sensing systems like traffic monitors) and (2) data collected for purposes other than its current use (such as cell phone or passenger flight data). The increasing volumes of data generated each day, especially the data referred to as "organic" (i.e., data collected for other purposes that are useful for social and behavioral science research), can help to solve today's complex societal challenges if they can be properly combined and interpreted while making use of new approaches and protecting the privacy of the individuals.

It is at this point that big data begins to affect the work of both the IC and the social and behavioral science communities. As social and behavioral scientists seek better and more varied kinds of data upon which to build evidence-based decision-making and understanding of human behavior, these big data trends impact their work as the ever-increasing amounts of data inevitably transform science, technology, economic activity, and other aspects of human behavior. Similar effects occur in the IC as it attempts to interpret and apply the data collected for or controlled by the IC in monitoring and containing threats to the nation's security and interests.

This common interest in big data and its social implications can serve as a foundation for the development of cooperative endeavors between the IC and the social and behavioral science communities. As the IC gears up to deal with the effects of big data and the increasingly social nature of U.S. security threats, it has a strong motivation to work with social and behavioral scientists and perhaps to make some IC-controlled data available to them to better understand the social, behavioral, and economic environment. The IC's desire to better engage with social and behavioral scientists in academia and industry, coupled with the social and behavioral scientists' desire for better and more varied opportunities, which may or may not include new sources of data, creates an opportunity for the potential cooperative endeavors this report explores.

Increasingly, social and behavioral science relies on the use of big data. Increasingly ubiquitous information and communication technologies and sensors of the physical environment, coupled with the ability to capture data from these systems and technologies nearly in real time, has produced a massive collection of growing datasets that document human behavior and measure the conditions and events in the physical environment. This large-scale data collection holds the promise of intelligent management of human interaction with the physical, social, and economic environment. It also presents tremendous challenges in managing and analyzing data, building models, generating theory, and, of course, ensuring privacy.

C. Workshop on Big Data for Social Sciences

The Big Data for Social Sciences workshop was held on May 11, 2012, to discuss the fundamental issues of applying big data to social science issues and methods for collaborating on big data research. The workshop provided the foundation for understanding how social and behavioral scientists view the application of big data to the issues they study. Selected findings from the workshop are as follows:

- Big data will change traditional quantitative social and behavioral science methodologies to include population-based observations in addition to experiments and sample surveys. New statistical approaches to weighting data to represent the population will be needed.
- Issues with big data that must be addressed include data acquisition, duration, and integration of datasets; ethical and legal considerations, including privacy concerns; and making data available to other researchers to replicate scientific results. Some solutions for sharing data with researchers are in place but these opportunities need to be expanded.

The agenda for the Big Data for Social Sciences workshop and a list of participants are included in Appendix A, and a summary of the workshop discussions is included in Appendix B.

D. Structure of the Report

The report first examines past and current IC and DOD programs and then explores how these programs could be adapted to begin building a community of practice focused on using social and behavioral science methodologies and theories to address IC questions. Where appropriate, a focus on the use of big data and smart cities is also explored.

The approaches to building a community of practice between the IC and social and behavioral scientists focuses on three goals examined in Chapter 2:

- Establishing a precedent
- Building a community
- Creating a funding source
- Creating visibility and vision

To understand how to accomplish these goals, MEDEA, DSSG, Minerva Research Initiative, and IARPA are discussed as examples of the IC and the DOD interacting with civilian science communities.

Building on these examples, Chapter 3 explores potential cooperative endeavors with social and behavioral scientists where the goal is to build a community of practice focused on establishing methods, tools, and datasets that help solve social and behavioral science issues. As appropriate, these suggestions are considered in the context of how big data can provide the basis for addressing social and behavioral science issues of common interest to both the intelligence and scientific communities. One example of a research topic, “smart cities,” is offered for discussion.

Chapter 4 provides a summary and conclusions.

2. Intelligence Community Interactions with Social and Behavioral Scientists

This chapter examines four necessary components to building a community of practice between the IC and social and behavioral scientists. First, it examines how to establish precedent for IC collaboration with a group of civilian scientists, using the MEDEA as an example. Second it discusses building a community of civilian scientists and introducing them to aspects of the Federal Government, using the DSSG as an example. Third, it discusses the establishment of continued funding for civilian scientists, examining both the Minerva Research Initiative and IARPA as possible models. These programs could provide a test bed for developing new tools, methods, and analytical findings for exploring the use of big data to study smart cities, which is a potential research topic of interest to both the IC and civilian scientists. In addition, social scientists can guide in designing the types of data collection and the questions that might be asked of data that are already collected routinely. Fourth, it discusses the role of a prominent advisory group to strategically and visibly plan the initiatives and provide credibility to the endeavors.

Exploring these programs can yield insight into the design of successful cooperative endeavors between the IC and scientists. Particularly informative are situations where big data plays a central role in the undertaking.

A. Establish a Precedent: Measurements of Earth Data for Environmental Analysis (MEDEA)

The IC established the Environmental Task Force in 1992, which became MEDEA in 1993.[2] MEDEA cleared about 60 environmental scientists for access to top secret documents with the mission to determine whether the scientific community studying environmental problems could use productively the monitoring assets of the IC.[3] In the process, and despite several concerns raised, the IC gained a measure of trust from civilian scientists.

---

2 Geoff Brumfiel, "Shared Intelligence," *Nature*, 2011, 477, 388−389.

3 Diane Snyder and Michael L. Brown, "The Intelligence Community and the Environment" In *Intelligence Reform in the Post-Cold War Era*, WWS 401a, 1997.

From the scientific community's perspective, MEDEA had a number of successes, including access in varying forms to relatively large quantities of previously classified data:

- Over 860,000 satellite photographs of the Earth taken between 1960 and 1972 were released[4]
- Hundreds of overhead images from six scientist-selected sites around the Arctic basin were captured and declassified[5]
- Satellite images of dozens of scientist-selected, ecologically sensitive areas around the world were captured[6]
- Images of areas of environmental interest for scientific research were archived at the Global Fiducials Library, among others[7]

In addition, the scientific community benefited from cooperation in data access, analysis, and interpretation as a result of MEDEA. Among the scientific advances attributed to MEDEA are archaeological discoveries about ancient civilizations, scientific advances in understanding the gradual thinning of Arctic ice, documentation and advances in the understanding of ecological changes concerning climate trends, increased knowledge of the movements of marine mammals, progress in detecting underwater seismic activity, and advances in weather prediction through greater understanding of wind patterns. Thus, MEDEA identified and made public data that were critically important to advances in environmental research. Such advances have the potential to spur future research that could yield even greater returns.[8,9]

Despite these successes, MEDEA raised a number of concerns from the civilian science community. First, the foundations of scientific research, "replicability—the ability of one scientist to reproduce another's findings using the same data—and verifiability—the ability to demonstrate the validity of the findings through

---

4 Brumfiel, "Shared Intelligence."

5 Committee on the Scientific Value of Arctic Sea Ice Imagery Derived Products; Committee on Climate, Energy, and National Security; National Research Council, *Scientific Value of Sea Ice Imagery Derived Products*, The National Academies Press, 2009, 48, ISBN 978-0-309-13763-8.

6 William J. Broad, "U.S. Will Deploy Its Spy Satellites on Nature Mission," *The New York Times*, November 27, 1995.

7 Brumfiel, "Shared intelligence."

8 Jeffrey T. Richelson, "Scientists in Black" *Scientific American*, February 1998, 48−55.

9 National Research Council, Committee on National Security Implications of Climate Change for U.S. Naval Forces, *National Security Implications of Climate Change for U.S. Naval Forces*, 2011, www.nap.edu/catalog.php?record_id=12914.

experimentation or observation,"[10] are threatened if data are available to only a small group of scientists, if the data collection method cannot be divulged, or if no scientist has access to equipment necessary to reproduce the data.[11] Though scientists will often embargo data and restrict access until they have completed their analyses, having the control of the data outside the scientific community may have made this issue particularly of concern for some scientists. Second, scientists must meet stringent requirements to access IC data, ranging from background investigations to obtain clearances to establishing secure laboratories and computer facilities for handling data. The requirements can be expensive in terms of time, administration, capital costs, and convenience.[12]

Projects such as MEDEA also raise concerns from the IC perspective. First, security is threatened as more people have knowledge of or access to the data and as more people know how data were collected and the capabilities of the collection process.[13] Second, while such activities may increase research quality as partners learn from each other, if the information they provide cannot be fully utilized, the IC may be distracted by working on problems that are not primary to its mission.[14] The longer term benefits of these cooperative situations can be lost in the concerns for short-term resources. Third, managing cooperative endeavors requires IC resources. IC staff must find good civilian partners and then monitor those partners to ensure that security requirements are met. Further, some effort must be spent in reviewing the classified material for declassification and in shaping secure data techniques that allow both partners to meet their needs.

In addition to these concerns, IC and civilian scientists come from different cultures. Many scientists strive to support open and free access to data.[15] In contrast, the mission of the Office of the Director of National Intelligence is "to gather and analyze the intelligence necessary to conduct foreign relations and national security activities."[16] For

---

[10] Richelson, "Scientists in Black." Scientific examples where replicability is not always possible, at least in the shorter run, include science around unusual events (tsunamis, meteorites, earthquakes, and volcanic eruptions) or the scientific information from the Mars rovers (which will not be reproduced for quite a while).

[11] The data from high-energy physics experiments involving thousands of people, few of whom have an overview, is another exception to these rules. See Steven Weinberg, "The Crisis of Big Science," *New York Times Review*, May 10, 2012.

[12] Snyder and Brown, "The Intelligence Community and the Environment."

[13] Ibid.

[14] Ibid.

[15] National Science Foundation, *Changing the Conduct of Science in an Information Age*, 2011, http://www.nsf.gov/pubs/2011/oise11003/.

[16] Office of the Director of National Intelligence website, http://www.dni.gov/index.php.

that reason, the IC is inherently biased toward releasing as little information as possible. Indeed, releasing data is likely to bring unknown or unexpected benefits to the IC regardless of the benefits to the civilian scientific community.[17]

For the MEDEA project, several critical conditions negated or minimized these concerns. First, the end of the cold war left a vacuum in the security requirements of the IC, and the IC was looking for initiatives that would solidify its reason for being and allow it to continue or extend operations.[18] Second, environmental scientists were eager to access the detailed big data from the IC and saw potential for scientific advances, because the IC systems had much greater capabilities for collecting data useful for environmental scientists than the civilian systems did at the time. Third, the United States generally supported the goals of environmental security. Though there were quibbles about exactly what the concept included, there was no organized opposition with sufficient strength to derail operations. Fourth, the costs were minimal and benefits to scientists high. The project involved a small fraction of the IC's entire security efforts. In many cases, the scientific advances came without substantial monetary costs to the nation, as the data were already in archives or the assets were made available to scientists during free time the assets had available. MEDEA was able to make a significant contribution to the nation and non-IC scientists through the release of archived data and the collection of new data for insignificant additional costs.

Because the concerns MEDEA raised were largely negated by the conditions just described, the IC also benefited from MEDEA by building trust among civilian scientists. The intelligence and scientific communities learned and benefited from each other. In particular, environmental scientists helped the IC with sensor calibration, validation, and applying state-of-the-art research techniques. Together they learned how to integrate the IC data with open information sources. Further, the collaboration created the visibility and excitement to attract new researchers to the scientific field and outside talent to the IC.[19]

These lessons learned are important if the IC does share data with the social and behavioral scientists. The engagement and value of the collaboration will dictate what they will do in the end. No one ever expected the declassification of data in the MEDEA project. No one knows what is valuable and to whom. Even if the IC does not share data, the major premise of this paper still holds that both the IC and social and behavioral scientists can

---

[17] Scott Pace, Kevin M. O'Connell, and Beth E. Lachman, "Using Intelligence Data for Environmental Needs—Balancing National Interests," Santa Monica, CA: RAND National Security Research Division, 1997.

[18] Robert Albro, "Minerva and Critical Public Engagement," The Social Science Research Council, The Minerva Controversy essays, 2008, http://essays.ssrc.org/minerva/2008/11/14/albro/.

[19] Pace, O'Connell, and Lachman, "Using Intelligence Data for Environmental Needs."

mutually benefit by collaborating on research projects. If the initial focus is on big data, accessibility to IC data may be less important.

B. Build a Community of Practice: Defense Science Study Group (DSSG)

To develop strong links between the national security community and emerging leaders in academia in science and technology disciplines, the Defense Advanced Research Projects Agency (DARPA) created the DSSG in 1985.[20] The original motivation for this program was the distrust held by many academics in the national security establishment due to the funding of the Strategic Defense Initiative. The DSSG provided a mechanism by which academic scholars could become more knowledgeable and involved objectively in national defense challenges. The program was aimed at a broad set of disciplines in science, engineering, and related disciplines.[21]

> The objectives of this 25-year-old program are to identify emerging leaders of S&T and introduce them to the national security community. The program is intended to instill an appreciation for the technical and operational challenges facing the national security community…and to foster in them a long-term interest in national security. Finally the program also seeks to create a network of informed and involved alumni, and to provide opportunities for those alumni to address national security concerns.[22]

Every other year about 15 candidates are selected for the 2-year program from about 200 outstanding nominees. The January 2012 *Report of the Defense Science Board Task Force on Basic Research* highlights the need for attracting the participation of the nation's top scientists and engineers in national security challenges. It singles out the DSSG as a unique program that has met that challenge for over 25 years. It recommends expansion of the program in order to increase the network of informed and engaged scientists and engineers exposed to the national defense community and its challenges. It further states that "this expansion should include an appropriate number of social and behavioral scientists, and medical researchers, insofar as those areas are among those chronically getting short shrift by DOD."[23]

---

[20] Richard J. Bergemann and Nancy P. Licato, *Summary Report of the Defense Science Study Group 1985–1988*, IDA Paper P-2310, September 1989, http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA217607.

[21] Interview with Robert Roberts, founder and coordinator of the DSSG program, September 17, 2012.

[22] Office of the Secretary of Defense for Acquisitions, Technology, and Logistics (ASDR&E), *Report of the Defense Science Board Task Force on Basic Research*, Washington, DC, January 2012.

[23] Ibid, p. 90.

One reason for DSSG's success is its continued support. After more than 25 years of operations, the DSSG showed impact by building knowledge, traditions, and trust. If it had run for only a few years, it would have had marginal to no effect. One of the primary successes has been to build a capacity of academic scientists with knowledge of technology and government activities that are able to provide input as researchers, government leaders, consultants, and over a career, a mix of these roles. Over the years, the DSSG has learned lessons critical to its smooth operations, such as those related to the amount of travel and timing of meetings. In the first year, academic researchers were selected for 3 years with nearly 30 days of travel and meetings per year. Today, DSSG members participate for 2 years and travel and meet for about 20 days per year, which appears to be optimal.[24]

C. Create a Funding Source: The Minerva Research Initiative and the Intelligence Advanced Research Projects Activity (IARPA)

The Minerva Research Initiative and IARPA offer two distinctly different examples of how agencies can elicit the involvement of civilian scientists in security-related projects. Differences in the structures and outcomes of these two funding programs provide critical insight into approaches for successfully garnering civilian scientist cooperation in IC and security-related issues.

1. Cautionary Tale: The Minerva Research Initiative

One of the more recent ventures into cooperation with the social and behavioral science communities has been the Minerva Research Initiative. Though Minerva was a DOD program, it involved security-related issues, so it may be informative for IC undertakings. Beginning in 2009, the DOD provided as much as $75 million over 5 years to fund social and behavioral science research in areas strategically important to U.S. national security.[25] The purpose was to increase DOD's knowledge of the applicable social and behavioral sciences, to expand the DOD's ability to manage social and behavioral science issues in a security context, to support social and behavioral science research in academia, to develop links between the DOD and social and behavioral scientists in academia, and to focus social and behavioral science expertise in academia on national security problems. One DOD researcher writes about the need:

---

[24] Interview with Robert Roberts, founder and coordinator of the DSSG program, September 17, 2012.

[25] Thomas Asher, "Making Sense of Minerva Controversy and the NSCC," Social Science Research Council, The Minerva Controversy essays. October 9, 2008. http://essays.ssrc.org/minerva/2008/10/09/asher/.

> [t]o understand the long term challenge posed by extremist ideology and what this means in nations experiencing rapid demographic changes. Language skills, cultural knowledge and understanding, understanding the attitudes of different populations, these are the critical tools that the US government needs to more fully integrate into our kitbag of capabilities for the future.[26]

Minerva Research Initiative funding was issued in two separate solicitations (1) a DOD Broad Area Announcement and (2) a separate announcement for funding through the National Science Foundation. The goal was to fund cross-institutional research centers and multidisciplinary research projects.

Unlike MEDEA and DSSG, the Minerva Research Initiative has drawn strong criticism from the academic community.[27, 28] Criticisms of the DOD's control of the original selection process led to criticism of the selection and funding procedures. The bulk of the criticism, however, focused on the content of the funded research. Earlier collaboration with social scientists may have changed the selection of and titles of the content.

Such criticism derives from concerns about subversion of social and behavioral sciences for malicious purposes in governance or military violence. These concerns have led to the creation of a set of essays posted by the Social Science Research Council under the title "The Minerva Controversy."[29] To date, commentaries reflect the opinions of supporters, detractors, and a few in the middle. Some concerns with social scientists developing working relationships with the military and intelligence communities focus on the possibility of scientists unintentionally participating in actions perceived to be supporting non-democratic actions or governments; the potential that research agendas might be influenced in a way that reduces academic freedom; the misrepresentation of research that could damage the reputation of social scientists as a whole; and a lack of openness of how results are used or misused.[30] Even using social scientists as consultants in an outsourcing

---

[26] Tom Mahnken, "Building Bridges and Communities," The Social Science Research Council, The Minerva Controversy essays, 2008, http://essays.ssrc.org/minerva/2008/12/30/mahnken/.

[27] James Petras, "Guest Editorial Procuring Academics for Empire: The Pentagon Minerva Research Initiative," *Dialect Anthropology* 33(1, March 2009): 1−4.

[28] Asher, "Making Sense of Minerva Controversy."

[29] The Minerva Controversy, a collection of essays sponsored by the Social Science Research Council, http://essays.ssrc.org/minerva/.

[30] Robert Albro, James Peacock, Carolyn Fluehr-Lobban, et.al. *Final Report on The Army's Human Terrain System Proof of Concept Program*. AAA Commission on the Engagement of Anthropology with the US Security and Intelligence Communities (CEAUSSIC), October 14, 2009. http://www.aaanet.org/cmtes/commissions/ceaussic/upload/ceaussic_hts_final_report.pdf; and Montgomery McFate and Steve Fondacaro, "Reflections on the Human Terrain System during the First 4 Years," National Defense University, *Prism* 4, No. 2, September 2011, http://www.ndu.edu/press/lib/images/prism2-4/Prism_63-82_McFate-Fondacaro.pdf.

way could lead to research being misused either intentionally or unintentionally by the military and intelligence communities, thus reducing accountability.

The topics in any open call for proposals must be broad enough to allow for a sufficient variety of research areas. If there is ambiguity about the program's ultimate purpose, however, critics may use that ambiguity to sow seeds of mistrust about the program. Past topics of the Minerva Research Initiative do little to dispel these criticisms. In its first year, the program sought proposals in five topic areas:[31]

1. Chinese Military and Technology Research and Archive Programs
2. Studies of the Strategic Impact of Religious and Cultural Changes within the Islamic World
3. Iraqi Perspectives Project
4. Studies of Terrorist Organizations and Ideologies
5. New Approaches to Understanding Dimensions of National Security, Conflict, and Cooperation

Further, when the funding source is the DOD (e.g., Minerva and DSSG) or the IC (e.g., MEDEA and IARPA), critics are likely to believe that the research serves the needs of the funding source rather than science. Thus, the restriction of topics to areas of clear DOD or IC interest and the funding source itself make it easy for critics to argue that the mission would do little to advance the science beyond the needs of the DOD or the IC. Assuming that argument to be credible, it would then be easy to paint any social or behavioral scientist working on the program as working in service of the DOD or the IC rather than serving science. This argument may be a bit simplistic in that all scientists must serve the focus of interest of their funders.

Part of the influence driving this argument among some social and behavioral scientists may be the limited prior experience they have had with DOD and IC work. When social scientists have done work for large corporations, they have often been able to publicize at least parts of the research and the research has been reasonably widespread among social and behavioral scientists. The security-related research has previously been less likely to be widely published and more often restricted to fewer researchers, so there may be less familiarity about the research among some in this community of scientists. If this lack of familiarity is the source of the criticism, the criticism should decrease as cooperation between scientists and the DOD or the IC increases; however, this increased cooperation needs to be proactively anticipated and managed.

[31] Asher, "Making Sense of Minerva Controversy."

The topics in the Minerva announcement for 2013 have evolved to focus on broader social and behavioral science issues:[32]

1. Belief Formation and Movements for Change
   a. Belief formation and influence
   b. Group identities and cultural norms
   c. Movements for change
   d. Collaboration and competition between violent groups
2. Models of Societal Resilience and Change
   a. Economic factors
   b. Energy, environment, and resource factors
   c. Other factors impacting societal stability and change
3. Theories of Power and Deterrence
   a. The role of the state in a globalized world
   b. Cyber norms and governance
   c. Beyond conventional deterrence
   d. Emerging topics in power and deterrence

Not all Minerva commentary from the civilian community has been negative. A moderate perspective in the controversy says that political scientists "have not been bought, as some crudely charge, nor have they been co-opted through more subtle means. They have retained their capacity for critical thinking; they have not become mere parrots of the official government line."[33] Though the Minerva Research Initiative came after the MEDEA project in time, it appeared to have been planned with little consideration of the MEDEA experience. Whereas MEDEA sought to declassify data with the goal to advance science, the Minerva initiative sought to have researchers conduct research to serve the DOD. The DOD did not work out the agreement so that social and behavioral scientists could be seen to benefit from the collaboration. This failure to recognize cultural differences may be at the root of many Minerva criticisms.

---

[32] For a history of Minerva, see http://minerva.dtic.mil/ and the recent Department of Navy, Office of Naval Research, Science and Technology, Broad Area Announcement #12-016, http://www.onr.navy.mil/~/media/Files/Funding-Announcements/BAA/2012/12-016.ashx.

[33] Robert R. Krebs, "Minerva: Unclipping the Owl's Wings," The Social Science Research Council, The Minerva Controversy, 2008, http://essays.ssrc.org/minerva/2008/11/19/krebs/.

One way to allay the Minerva funding concerns is to invest in infrastructure: language training, pedagogical and research materials, training grants, travel opportunities, and the like. Rather than dictating major topics for research, it would have scholars themselves select projects that would deepen understanding of regions and trends that impinge on DOD operations.[34]

2. Potential Model: IARPA

The IARPA was created in 2008 to invest "in high-risk, high-payoff research programs that have the potential to provide the United States with an overwhelming intelligence advantage over future adversaries."[35] IARPA tackles the difficult research challenges that span across or fill gaps between intelligence agencies, while the other 16 agencies in the IC focus on the day-to-day concerns within their own particular sphere of operation. The primary expectation is for IARPA to create transformational solutions that can be transitioned to the IC, but spillover to the civilian population is expected. With no operational mission and no facilities to deploy technologies directly to the field, IARPA operates on an intermediate to long-term timeline, seeking to address future intelligence problems by creating solutions that IC agencies can implement before the problems become critical.[36]

IARPA uses the DARPA model of providing funding to academia and the private sector to achieve scientific breakthroughs for IC use. Programs are typically funded for set durations of 3 to 5 years with clearly defined and measurable goals. Funds are administered by three offices:[37]

- The Office of Smart Collection funds programs to improve the information search paradigms and to enhance the value of gathered information to reach the appropriate users in a timely manner. The Office of Incisive Analysis funds programs to extract maximum insight from collected information.
- The Office of Safe and Secure Operations funds programs to create solutions to counter the capabilities of adversaries that might inhibit free and effective U.S. operation in a networked world.

[34] David Engerman, "Knowing the Enemy," The Minerva Controversy, The Social Science Research Council, 2009, http://essays.ssrc.org/minerva/2009/01/23/engerman/.

[35] IARPA – Our Organization, http://www.iarpa.gov/organization.html.

[36] Ibid.

[37] Ibid.

These offices' programs have funded researchers from a spectrum of disciplines, with much of the research involving various forms of big data. IARPA publicly acknowledges programs to develop the following:

- New concepts for advanced signal sensing and processing[38]
- Tools to identify novel information sources; extract useful intelligence; secure communication; and improve tagging, tracking, and location techniques[39]
- New technologies that dramatically extend the operational capabilities of surveillance and reconnaissance unmanned aerial vehicles[40]
- Computational techniques to analyze and securely manage large data flows[41]
- Techniques to recognize, categorize, and interpret metaphors in plain text for different cultural perspectives[42]
- Methods for automated, continuous analysis of publicly available data including web search queries, blogs, Internet traffic, traffic webcams, and financial markets to detect or anticipate societal events[43]

IARPA has been in existence for too little time to adequately assess its success; however, some positive conditions and characteristics that bode well for its success include:

- Major, stable funding
- Flexibility in addressing a changing variety of problems
- Anticipated spillovers into civilian and commercial applications
- Headquarters location at a university signals a desire to partner with academia and industry
- Problems (often rooted in big data) that are attractive to a broad group of scientists beyond the IC interest and issues

One challenge for IARPA is that it uses many different datasets and technologies. Public awareness of certain funded projects or programs could raise privacy concerns and

---

[38] HFGeo Program - Broad Agency Announcement (BAA).

[39] Office of Smart Collection Office Wide - Broad Agency Announcement (BAA).

[40] Great Horned Owl (GHO) Program - Broad Agency Announcement (BAA).

[41] Office of Safe and Secure Operations Office Wide - Broad Agency Announcement (BAA).

[42] Metaphor Program - Broad Agency Announcement (BAA)..

[43] Open Source Indicators (OSI) Program - Broad Agency Announcement (BAA).

taint IARPA's public image. And assessing tradeoffs of privacy with increased benefits to society is an area that requires immediate attention.

Another related model that could work is the NIST Advanced Technology Program[44] in which topics and approaches are proposed by the community but that must be relevant to both the IC; it is a bottom-up approach, not a top-down approach. Broad areas of research and specific criteria are critical and then ideas are generated by the community.

D. Establish an Advisory Board or Task Force

There are many examples of IC and DOD scientific advisory boards that provide independent advice on science and technology matters relating to the mission of the organization that they serve. The members are usually well-known experts who provide vision, credibility, and guidance. For example, the Air Force Scientific Advisory Board (SAB) is described as follows:

> The SAB is a Federal Advisory Committee that provides independent advice on matters of science and technology relating to the Air Force mission, reporting directly to the Secretary of the Air Force and Chief of Staff of the Air Force. Since 1944, luminaries such as Dr. Theodore von Kármán, General James Harold "Jimmy" Doolittle, Dr. Ivan Getting, Dr. Edward Teller, and Dr. Charles Stark Draper, and more recently Dr. Robert Lucky and Mrs. Natalie Crawford, have provided visionary and forward-looking advice on technologies such as: supersonic aircraft, weather forecasting, satellite communications, medical research, crewless airplanes, and defenses against aircraft and missiles.[45]

The Defense Science Board (DSB), which is composed of well-known experts in the defense science and technology world, is known for its role in producing reports that are often future-oriented as well as advisory. The DSB's membership is composed of well-known experts in the defense science and technology world. Its most recent newsletter provides several examples of these reports. For example, the 2012 study titled "Technology & Innovation Enablers for Superiority in 2030," and focuses on emerging technologies that will enable the next generation of dominant military capabilities to be in development or fielded by 2030.[46]

---

[44] See About the ATP (Advanced Technology Program) at http://www.atp.nist.gov/atp/overview.htm.

[45] See About the SAB at http://www.sab.af.mil/index.asp.

[46] For other examples of DSB projects, see the May 2012 DSB newsletter: http://www.acq.osd.mil/dsb/newsletters/2012-05-DSB_Newsletter.pdf.

The Intelligence Science Advisory Board was replaced in 2010 by a Senior Advisory Group that convenes task forces on specific issues as needed.[47] Similar to a scientific advisory board, a task force structure could be established to guide and oversee collaboration between the the IC and the social and behavioral and social science community. Importantly, the choice of the task force members, their prominence, and their ability to create excitement and credibility in the communities are the critical components of such a task force.

The proposed IC Social and Behavioral Science Advisory Board (or task force) should be composed of IC experts and prominent computational statisticians, social and behavioral scientists, and physical scientists.

E. Summary and Lessons for Future Collaborations

This chapter discusses three necessary components to interacting with a civilian science group, establishing a precedent, creating a community, and creating a funding source using four programs as examples where civilian scientists cooperated with the IC or DOD. Two of the programs, MEDEA and DSSG, have been clearly successful in some major aspects. The Minerva Research Initiative has been subjected to intense criticism, and IARPA is too new to fully assess. As a group, these programs suggest clear lessons for the development of cooperative endeavors between the IC and the social and behavioral science communities.

Table 2 summarizes the positive outcomes of the example programs as they relate to as the three necessary components. These outcomes underlie the approaches proposed in the next chapter for the IC to build successful and enduring cooperative endeavors with social and behavioral scientists.

---

[47] See the Federation of American Scientists (FAS), Secrecy News, "DNI Disbands the Intelligence Science Board," October 14, 2010, http://blogs.fas.org/secrecy/2010/10/isb_disband/.

Table 2. Summary of Positive Outcomes Tied to Necessary Components for Building a Community of Practice

Necessary Components	Example Programs	Outcomes
Establish a precedent	Measurements of Earth Data for Environmental Analysis (MEDEA)	Released useful data to civilian scientists Built trust between IC and civilian scientists
Build a community	Defense Science Study Group (DSSG)	Established group of civilian scientists with IC interests Introduced group to Federal Government Used group to solve IC problems
Create a funding source	Intelligence Advanced Research Projects Activity (IARPA)	Funded projects with IC and civilian scientist interest
Create an advisory board or task force	Air Force Scientific Advisory Board (SAB)	Provides visibility, credibility, and guidance for collaborative efforts

3. Proposed Approaches for IC Cooperation with Social and Behavioral Scientists

This chapter explores four potential approaches to developing cooperative endeavors between the IC and the social and behavioral science communities:

1. Creation of a MEDEA-like program for the social and behavioral sciences
2. Creation of an IC-sponsored Social and Behavioral Sciences Study Group like the successful DSSG
3. Expansion of IARPA to fund the social and behavioral sciences or creation of a new funding program that allows for the community to propose ideas within broad areas of research and criteria.
4. Examination of "smart" cities as a test bed for IC cooperation with social and behavioral scientists

Programs related to MEDEA and IARPA would focus more broadly on social and behavioral science issues that rely on big data because these issues are of growing interest to the IC and social and behavioral scientists. In the current digital environment, data are a crucial element of any program that seeks to garner cooperation from civilian scientists, specifically what data can be studied and what freedom the scientists have in exploring the data and publishing their findings. These aspects played important roles in the MEDEA and IARPA examples and underlie the approaches presented here, whether explicitly mentioned or not. The criticality of big data to these approaches stem from:

- *The potential for big data endeavors to extend current methodological tools and scientific findings.* Researchers also recognize that companies now need "data scientists" who have skills to analyze big data, so they are attracted to these projects as well.[48]

[48] Thomas H. Davenport and D. J. Patil, "Data Scientist: The Sexiest Job of the 21st Century," *Harvard Business Review* (October 2012, reprint R1210D): 8.

- *The usefulness of big data to detect, analyze, and predict patterns necessary to understand and predict patterns of human behavior.* The success of these projects often hinges on acquiring and analyzing large quantities of data.[49]
- *The role of big data in automated management.*[50] The adage that "you can't manage what you don't measure" will soon become "you can't automate management of what you don't measure." As the digital technology opens greater opportunities for automated management, the IC must work with researchers to get ahead of the curve or risk losing data collection capabilities.

The success of any program that attempts to build a community of practice between IC and social and behavioral scientists will lie in the details of the implementation. For the programs recommended in the following sections, big data is an important detail. Unfortunately it is a detail that depends upon the specific situation and technology. Thus, the data for these programs cannot be designated ahead of time; the importance of the data can only be highlighted.

A. Establish a Precedent for Interaction

A program built on the MEDEA model could begin with selection of a group of outstanding social and behavioral scientists from academia and industry who meet top secret clearance requirements.[51] These civilian scientists could meet with IC representatives to ascertain jointly what IC data could be released or what assets could be redirected to provide data that supports social and behavioral science research.[52] The results of these meetings could be communicated through both IC reports and published research papers. The program characteristics and project topics of a MEDEA counterpart would have to address the conditions previously mentioned as supporting the success of MEDEA:

1. *Clarity and strength of support for the IC policies and initiatives.* The IC now has a clear, strongly supported set of policies and initiatives. Projects that

---

[49] Dominic Barton and David Court, "Making Advanced Analytics Work for You," *Harvard Business Review* (October 2012 reprint R1210E): 7.

[50] Andrew McAfee and Erik Brynjolfsson, "Big Data: The Management Revolution," *Harvard Review* (October 2012, reprint R1210c_: 9.

[51] Even if the community were to work on big data issues using unclassified data, it may still be a good idea for them to have top secret clearances to better understand the IC analytical needs.

[52] Other government agencies and scientific communities have faced and dealt with these concerns. For example, see Hal Abelson and Lawrence Lessig, "Digital Identity in Cyberspace," White Paper submitted for 6.805/Law of Cyberspace: Social Protocols (December, 10, 1998), http://groups.csail.mit.edu/mac/classes/6.805/student-papers/fall98-papers/identity/linked-white-paper.html.

require extension of the current IC agenda seem outside the realm of consideration if they require bipartisan support and a budget increase unless the projects could be positioned and promoted as remedying an ongoing emergency situation. The MEDEA counterpart would have to demonstrate a strong fit with current IC policies and initiatives or at least be neutral with respect to those policies and initiatives.[53]

2. *Eagerness for the data in social and behavioral science communities.* As the Minerva Research Initiative demonstrated, working with the DOD or IC is often not considered a reputation builder within the civilian social and behavioral science communities. The perceived risk to the social or behavioral scientists' reputations could be offset by the opportunity to work with new data or on new topics with the clear potential to contribute to dramatic advances.
3. *Broad civilian support for the goals of the project.* Project areas where there is strong civilian support nationwide or at least weak civilian resistance or criticism would be good candidates for a MEDEA counterpart. Topics with strong civilian support might include the social, behavioral, and economic issues relevant to promoting the economy and jobs or deterring major crimes.[54]
4. *Minimal costs.* In the cases where data are provided, a program that redirected selected IC assets to address societal issues could claim minimal costs and leverage previous investments.
5. *Mutual benefit and learning.* A program where both the IC and the civilian scientists learned from one another would provide mutual benefit and reliance and foster trust.

Developing a MEDEA counterpart for the social and behavioral sciences is not an easy task. A multitude of pitfalls could impair the success of a newly established program, even if the program merely contributes data and leaves the analysis to civilian scientists as with MEDEA.[55] Much of a new program's risk is at the public perception level, including the perceptions of civilian social and behavioral scientists and public citizens. For that reason, any attempt by the IC to implement this type of program should be preceded by research and outreach and input from prominent experts.

---

[53] Another example is the emphasis on "human terrain" in the Afghan and Iraqi engagements. See Albro, Peacock, Fluehr-Lobban, et al., *Final Report on the Army's Human Terrain System.* Also see McFate and Fondacaro, "Reflections on the Human Terrain System."

[54] These topics are perhaps not of direct relevance to the IC, but would allow a platform for developing new tools and approaches.

[55] Snyder and Brown, "The Intelligence Community and the Environment."

B. Build a Community of Social and Behavioral Scientists

One option for building a community would be to expand the DSSG to include more social and behavioral scientists by selecting physical scientists the first year and social and behavioral scientists the next. This option would rely on processes already in place but require selection every year rather than every other year. Some activities between the two groups could overlap so each would benefit from each other's disciplines. However, an internal IDA group conducted and reviewed the DSSG and its potential applicability to the social and behavioral science communities. The group concluded that the social and behavioral scientists could not be an "add on" to the physical scientists for several reasons. First, the physical science community has been directly involved in defense research for many decades. Second, social and behavioral scientists study different topics, and the two communities have distinctly different cultures. Third, many of the visits arranged for physical scientists would not be of particular interest to the social or behavioral scientist, nor would they highlight the areas in which social and behavioral sciences might play.[56]

A second approach would be the creation of an IC-sponsored group that brings together a mix of scientists (including social, behavioral, physical, biological, and computational) from academia and elsewhere to meet with government agencies, intelligence agencies, and Congress and work together on issues related to the role of intelligence activities in societies, lifestyles, economies, and government. A mix of scientists allows sharing of knowledge, assures a better understanding of issues from multiple perspectives, and provides solutions that better integrate multiple disciplines. Like they do in the DSSG, scientists might serve in an active role for a modest number of years and transition to alumni where they would be available for participation and consultation as needed, although there could be more inducements to keep the alumni group more active, such as setting up regular grants or funding programs that would encourage research on specific topics or creation of methods and tools.

This group might be funded by IARPA, analogous to DARPA's funding of DSSG today. This program would focus on attracting the best and brightest social and behavioral scientists to get involved on behalf of the IC. As this approach has done with the DSSG and DOD, this could lead to a longstanding cadre of social and behavioral scientists who apply their expertise to IC issues. One model would be to engage graduates of the program on a select basis as was done by the successful MEDEA

[56] S. K. Numrich, IDA, personal communication, November 5, 2012.

program. One additional benefit of a dedicated program is that the stigma of supporting the IC could potentially be removed throughout the educational process.[57]

A program along the lines of the DSSG for intelligence scientists would build alliances with outstanding social and behavioral scientists, create networks for knowledge sharing between disciplines, and increase IC access to the services and support of these scientists. Additionally the program could demonstrate that the IC and civilian scientists from all disciplines can work to achieve positive goals. Also, select members of the intelligence group and the scientist group could be brought together for joint activities that would benefit the two groups specifically and the IC and the DOD more broadly.

C. Create a Funding Source by Expanding IARPA to Study Big Data

One option would be to expand the IARPA program to fund research that uses big data to solve social and behavioral science problems. The research could address social and behavioral science problems of the proposers' choosing, as long as the research was interdisciplinary and had the potential to reasonably provide novel practical insight into complex socio-behavioral problems. If desired, the initial projects might also be restricted to having some big data component. In addition to progress reports and updates, the deliverables on the research could include a final report that demonstrated the advancements in knowledge and behavioral insight or a report that applied the insights to a relevant scenario constructed by the IC sponsor. Other outputs could include implementation of a new system or approach to delivering services within a city, university, or government. These new systems would be driven by the use of real-time data and bringing together multiple sources of data.

This program would benefit social and behavioral scientists by funding a broad area of research, funding doctoral and postdoctoral students, and providing the opportunity to work on interdisciplinary projects that demonstrate the practical applicability of social and behavioral science theories. Benefits to the IC would include the development of a relationship with outstanding social and behavioral scientists and the potential to have those scientists apply their work to generalized and declassified versions of IC problems.

The major disadvantage is that funding this program could be relatively expensive. Costs would include the cost to fund the projects and administer more grants. If projects were selected well, however, this approach could provide practical returns in just a few years. The goal is to work on high-risk projects in which a few projects are expected to produce large benefits.

---

[57] Personal communication with Robert Roberts, founder and coordinator of the DSSG program, November 5, 2012.

D. Examine Cities as Test Beds for IC Cooperation with Social and Behavioral Scientists

Smart cities have sparked interest around the world to improve urban efficiency, resilience, and sustainability. In 2011, a conceptual paper on smart cities listed over 75 smart cities worldwide.[58] Some of the smart city projects, such as those in London and New York, evolved from the networking of buildings, research parks, and financial centers as the cities and businesses tried to exploit already existing competitive advantages.[59] Others, such as Singapore, tried to gain competitive advantage by building cutting-edge information and communication technology to transform into an information economy.[60, 61, 62]

Thus, a fourth approach for the IC's cooperation with social and behavioral scientists would be the development of new methods and tools to leverage a variety of "smart cities" in select U.S. locations and to better understand such cities overseas. This approach requires more discussion than the approaches discussed thus far to allow more consideration of the underlying big data.

While the smart city concept has been defined in a variety of ways,[63] this report considers the smart city to be an area of urban innovation where the planners, developers, managers, and inhabitants strive to create, in a densely populated environment, innovative governance, efficient infrastructure management, economic growth, and enriching, sustainable lifestyles.

To accomplish these goals, the smart city enterprise must be technologically equipped with the abilities to (1) collect observational data from the infrastructure, environment, and people; (2) get the data to the right place at the right time; and (3) integrate, analyze, and act upon data collected from diverse sources.

---

58 Taewoo Nam and Theresa A. Pardo, "Conceptualizing Smart City with Dimensions of Technology, People, and Institutions," In *The Proceedings of the 12th Annual International Conference on Digital Government Research*, Dg.o'11, June 12−15, 2011, College Park, MD.

59 John G. Jung, "Smart Communities: Digitally-Inclined and Content-Rich," *New Telecom Quarterly* (1Q, 1998): 19–26.

60 Mahmoud Al-Hader and Ahmad Rodzi, "The Smart City Infrastructure Development & Monitoring," *Theoretical and Empirical Researches in Urban Management*, 2 (11, 2009), 87−94.

61 Warwick Neville, "Managing the Smart City-State: Singapore Approaches the 21st Century," *New Zealand Geographer* 55 (1, 1999,): 35−45.

62 Bob Hodgson, "Technopolis: Challenges and Issues," In David V. Gibson, George Kozmetsky, and Raymond W. Smilor (eds.), *The Technopolis Phenomenon: Smart Cities, Fast Systems, Global Networks*, Lanham, MD: Rowman & Littlefield Publishers, Inc., 1992.

63 Nam and Pardo, "Conceptualizing Smart City."

By providing data and collaborating with a variety of smart city test beds across the nation, the IC would be able to contribute skills, data, and professional experience in areas where its expertise clearly surpasses the civilian community. The goals of the social and behavioral scientists and the IC are different here. The IC could be trained by learning about and using smart city data and it might also spill over to other areas of IC interest. Social and behavioral scientists can apply what they learn to developing and enhancing smart city life.

As with any project, the smart city is only as good as its planning and execution or management. In Singapore, emphasis was on the planning and development of the digital infrastructure.[64, 65, 66, 67] Singapore and other early smart cities focused on building a cutting-edge infrastructure that would serve as an operational platform for the efficiency "killer app" expected to come along later and allow these cities to operate more successfully than other cities because of their more efficient high-technology infrastructure. However, the "killer app" for cities may more likely be "sociability, not efficiency"[68] and other dimensions of smart cities, including resilience, sustainability, quality of life, equity, and engagement.

Table 3 presents the general processes involved in managing the technological hub at the heart of smart cities. The processes begin with the collection and transmission of observational data throughout the city's infrastructure. Data streams include fixed and mobile sensors; mechanical and human observations; tabular, web, and transactional data; and other outputs through current and future collection methods.

These data create the foundation of the smart city. The management structures and processes are built from the interactions and collaborations that occur after the data are collected. In the ideal smart city, collaborations occur on different levels with people with different skill sets and professions. Typically mentioned are collaborations that integrate the skills of data managers, statisticians, economists, social and behavioral scientists, urban planners, city managers, policy makers, service providers, and technology and application developers. Also important are the interactions and collaborations with and among the inhabitants.

---

[64] Al-Hader and Rodzi, "Smart City Infrastructure Development & Monitoring."

[65] Jung, "Smart Communities."

[66] Arun Mahizhnan, "Smart Cities—The Singapore Case." *Cities* 16 (1, 1999), 13−18.

[67] Sam Allwinkle and Peter Cruickshank, "Creating Smarter Cities: An Overview," *Journal of Urban Technology*18 (2, 2011): 1–16.

[68] Carlo Ratti and Anthony Townsend, "The Social Nexus," *Scientific American* 305 (2011): 42–48.

Table 3. Social, Behavioral, and Technical Processes in Smart City Management

Process	Description
Sensing	Collecting observational data in formats that allow integration and aggregation
Transmitting	Sending and receiving information between and among systems and people in private and public formats with single and multiple recipients
Analyzing	Extracting and interpreting the status of observed attributes and the relationships among data items
Modeling	Communicating the significance of data relationships through demonstrations and predictions of social, behavioral, and economic impact
Collaborating	Discussing, assessing, valuing, and promoting the social, behavioral, and economic impact of system statuses and proposed changes
Intervening	Instituting changes and studying feedback

Table 4 describes the three fundamental components that form the core distinctions among smart cities. Different smart cities gave different emphasis to these components as they developed their version of the smart city. Those implementations varied with the sources and levels of funding, the specific technology companies involved in infrastructure development, the workforce and commercial compositions of the area, and the culture and management style of the organizations leading the implementation.

Table 4. Core Components of the Smart City

Component	Description
Technology	Characteristics of the information and communication technology involved in the sensing and transmission of messages and broadband content
Institutional	Characteristics of the organizations and policies involved in the planning, governance, regulation, and service provision for the jurisdictional area
Human	Characteristics and diversity of the entrepreneurs and workforce, the social, education and cultural environments, and the creative outlets

Source: Taewoo Nam and Theresa A. Pardo, "Conceptualizing Smart City."

As a city-state, the central government in Singapore plays a dominant role in all urban affairs. It is uniquely positioned to develop an implementation plan wherein institutional and technological factors dominated planning efforts. In building the smart city, the central government decided what infrastructure would be installed, how it would be used to provide services and extract efficiencies, and how the people would adapt to that infrastructure. Traffic was managed by taxing vehicle ownership, pricing use of the roads, and mandating that vehicles have cash card technology to allow automated electronic charges.[69, 70] The population was largely relocated to high-rise residential buildings with quotas on ethnic concentrations to manage racial tensions and allow for more efficient infrastructure access.[71]

Few cities have political and cultural conditions that allow an approach as extreme as the one used in Singapore. Though cities may increasingly emphasize smart technologies, they may do so with a more balanced application of technological, institutional, and cultural factors. Cities such as New York, London, and Paris are finding that opening formerly private data to the public empowers entrepreneurs not only to better meet citizens' needs but also to compliment the function of many public institutions.[72]

Smart cities have already demonstrated a number of societal benefits, including reducing the consumption of water and energy, reducing traditional urban construction, improving transportation, providing more efficient and productive interaction with services, and allowing city managers to smooth the demands for energy, water, and transportation.[73] The field is still young and has only begun to explore the ways that more open access to the infrastructure, data, institutions, and inhabitants can impact urban life.

In addition to smart cities creating living experiments for the social and behavioral sciences, big data coming out of smart city projects has the potential to create paradigm shifts that may advance social and behavioral sciences theories. The data will open areas that social and behavioral scientists have been unable to explore because their traditions require either high-purity data collected by surveys numbering in the thousands or statistical data flows that sample widely but thinly, offering little opportunity to meaningfully parse and segment the data. Where these methodologies are non-real time and expensive, the data from smart cities will be real time, extensive, and comparatively cheap but will require a different set of algorithms and processes for analyses. Developing these algorithms will be complicated, requiring multiple approaches, but the situation will improve as researchers gain experience.

In summary, smart cities offer an opportunity for the IC to do the following:

- Contribute technology and expertise in an arena where its capabilities exceed the civilian community

---

[69] Mahizhnan, "Smart Cities—The Singapore Case."

[70] Neville, "Managing the Smart City-State."

[71] Ibid.

[72] Ratti and Townsend, "The Social Nexus."

[73] Colin Harrison and Ian Abbott Donnelly, "A Theory of Smart Cities," in *Proceedings of the 55th Annual Meeting of the International Society for Social Science*, July 17–22, 2011, University of Hull Business School, UK.

- Develop capabilities and technology compatibilities with the population centers of the future
- Assist in setting national standards for technology integration and data storage and transmission
- Partner with social and behavioral scientists eager to participate in the projects
- Collaborate in an environment with mutual benefits for the IC, social and behavioral scientists, and the civilian community

4. Summary and Conclusions

A community of practice between the U.S. Intelligence Community (IC) and social and behavioral scientists is proposed for the purpose of addressing issues of importance to the IC, researching topics that are both relevant to social and behavioral scientists and of mutual benefit to that community and the IC, and focusing attention on the intelligent management and use of big data.

At a time when the government, the private sector, and others are working to solve the challenges of handling big data, the IC and social and behavioral scientists have several opportunities to cooperate and collaborate productively in ways that meet the needs and requirements of both the intelligence and scientific communities. Two cooperative opportunities are inherent in the deployment of technology-based infrastructures and the trends in implementing automated systems for smart cities. For these cooperative endeavors to succeed, it will be necessary to establish guidelines to ensure that the programs address the issue of privacy and that they are positioned positively to the public. In addition, the programs must provide are obvious benefits and protections to the public, the social and behavioral sciences, *and* the IC.

This report describes approaches for new initiatives based on an analysis of previous cooperative programs. It begins with a discussion of the motivation for selected IC-related (and DOD-related) cooperative undertakings with scientists outside the intelligence and military community and then explores the options and concerns in IC cooperative endeavors with social, behavioral, and economic scientists to address social issues. The report then examines four approaches that the IC could explore in developing cooperation to build a community of practice with a focus on establishing methods, tools, and datasets that solve social and behavioral science problems. The approaches are:

- Creating a social, behavioral, and economic scientist-based counterpart to the MEDEA community. This would allow social and behavioral scientists to have access to unique sources of large data.
- Sponsoring an Intelligence Social and Behavioral Sciences Study Group modeled on the DSSG, or expanding the current program to include more social and behavioral scientists either in the current program or by creating a separate social and behavioral scientist cohort, selected in the alternative year to the current DSSG cohort. Such a program would build trust in social and behavioral science communities.

- Expanding IARPA to fund social and behavioral scientists to conduct research on social and behavioral science issues that require big data, and considering a Minerva Research Initiative-like program but avoid the pitfalls of the current program. IARPA and a Minerva-like program would provide funding to the academic communities.
- Applying IC technologies and methodologies to support the development of a variety of “smart” or “intelligent” cities in selected locations that will further advance uses of big data as well as understanding the cities of the future.

The next steps could be to bring together the IC and social and behavioral scientists to begin to plan how collaborations between them would work. To give the initiative visibility and to obtain buy-in, one suggestion is to create an advisory group of well-known and respected social and behavioral scientists to participate in establishing a governance structure and approach for the IC and social scientists to collaborate.

Appendix A.
Big Data for Social Sciences
Workshop Agenda and Participants

Friday May 11, 2012

Big Data for Social Sciences Workshop Agenda

8:00–8:30	Continental breakfast
8:30–8:45	**Introductions**
9:00–9:30	**Workshop Goals - Setting the Stage** • What critical social science questions can big data answer? • How is "big data" defined—what does it include? Who are the users?
9:30–11:15	**Data Collection, Integration and Exploitation** • Data streams and sensing technologies for big data • Proprietary and privacy considerations • New techniques for data collection, data integration, and data exploitation
11:15–12:00	**Lunch & Introduction to Case Studies, (Smart Cities, Crime and Social Problems, and Civic Engagement)**
12:00–1:30	**Case Study Breakout Groups:** • Smart Cities • Crime and Social Problems • Civic Engagement
1:30–2:00	**Case Study Report Backs**
2:00–2:45	**Collaborations** • What effective collaborations exist with respect to data collection, integration, and exploitation? • How can we build collaborations that support interdisciplinary challenges? • What incentive structures may exist to encourage interdisciplinary collaboration aimed at solving complex social issues?
2:45–3:00	**Wrap-up**

Breakout Groups

Smart Cities	Crime and Social Problems	Civic Engagement
Bhaduri, Budhendra	Brady, Henry	Ansolabehere, Stephen
Eardley, Doug	Earl, Jennifer	Fienberg, Stephen
Eddy, William	Gutmann, Myron	Imam, Neena
Silva, Claudio	Land, Ken	Koonin, Steve
Wang, Shaowen	Turner, Margery	
Facilitators:		
Shipp, Stephanie	Parker, Rachel	Walejko, Gina

Smart Cities (land use, consumption, transportation, and urban planning)

1. What are key components to designing and living in Smart Cities?
2. What are interesting research questions unique to smart cities?
3. What data do we need to answer these research questions? What data exist? What data don't we have, and why don't we have it?
4. What new data streams or sensing technologies need to be developed or repurposed to answer these research questions?
5. What are the challenges to using big data to answer these research questions?

Crime and Social Problems (violence, civil unrest, social control, and policing)

6. What are interesting research questions unique to crime and social problems?
    a. What are some potential social implications of studying crime and social problems using big data?
7. What data do we need to answer these research questions? What data exist? What data don't we have, and why don't we have it?
8. What new data streams or sensing technologies need to be developed or repurposed to answer these research questions?
9. What are the challenges to using big data to answer these research questions?

Civic Engagement (voting and elections, community development, and public participation)

10. What are interesting research questions that are unique to civic engagement?
    a. What are the social implications of studying civic engagement using big data?
11. What data do we need to answer these research questions? What data exist? What data don't we have, and why don't we have it?
12. What new data streams or sensing technologies need to be developed or repurposed to answer these research questions?
13. What are the challenges to using big data to answer these research questions?

Attendees

Stephen Ansolabehere, *Harvard University*

Budhendra Bhaduri, *Oak Ridge National Laboratory*

Henry E Brady, *University of California, Berkeley*

Douglas Eardley, *University of California, Santa Barbara*

Jennifer Earl, *University of Arizona*

William F Eddy, *Carnegie Mellon University*

Stephen E Fienberg, *Carnegie Mellon University*

Myron P Gutmann, *National Science Foundation*

Neena Imam, *Oak Ridge National Laboratory*

Kenneth Land, *Duke University*

Claudio T Silva, *Polytechnic Institute of New York University*

Margery Turner, *Urban Institute*

Shaowen Wang, *Helen Corley Petit Scholar of the College of Liberal Arts and Sciences*

Workshop Organizers:

Steven E. Koonin, *IDA Science and Technology Policy Institute and New York University Center for Urban Science and Progress*

Sallie Ann Keller, *former Director, IDA Science and Technology Policy Institute and current Vice-President Academics & Provost, Professor of Statistics, University of Waterloo*

Rachel Parker, *IDA Science and Technology Policy Institute*

Stephanie S. Shipp, *IDA Science and Technology Policy Institute*

Gina K. Walejko, *IDA Science and Technology Policy Institute*

Tom Milani, *IDA Studies and Analysis Center*

Appendix B.
Workshop Summary

Workshop Setting

IDA researchers organized and hosted a workshop called Big Data for the Social Sciences in Washington, DC, on May 11, 2012. Over the course of the day, a group of 17 academic and nonprofit researchers, along with several IDA research staff members gathered to exchange ideas on these topics.

The workshop began with introductions and a statement of goals. A moderated discussion of data collection, integration, and exploitation followed. Next, breakout groups were assigned three topics: smart cities, crime and social problems, and civic engagement. The groups assembled, discussed their topics, and determined the key points. The entire group reconvened, and a member of each breakout group presented the group's findings. The workshop concluded with a discussion of collaborations, followed by a wrap-up and recommendations for next steps.

Workshop Motivation

The world outside the Intelligence Community (IC) is rapidly building capabilities in computational social science and statistics. Financial companies, retailers, web and mobile app developers, among other commercial enterprises, are collecting and analyzing large scale datasets, while public and private sensor equipment and video surveillance tools are proliferating. Further, social media has become a trove of textual, video, and photo data. At the same time, applied behavioral and social science research techniques used in the private sector raise questions about data ownership on the one hand, and privacy, on the other.

To capitalize on, and stay abreast of these trends, the IC, among other stakeholders, is seeking ways to apply new and existing data, methods of data collection and analysis capabilities to broader societal challenges; to develop, refine, and validate methodologies in an unclassified setting; and to train and engage a new generation of computational social scientists. As one part of a task for the IC, IDA conducted a workshop bringing together experts from multiple disciplines to focus on the possibilities of using large datasets to address issues studied by and of concern to social scientists.

Workshop Summary

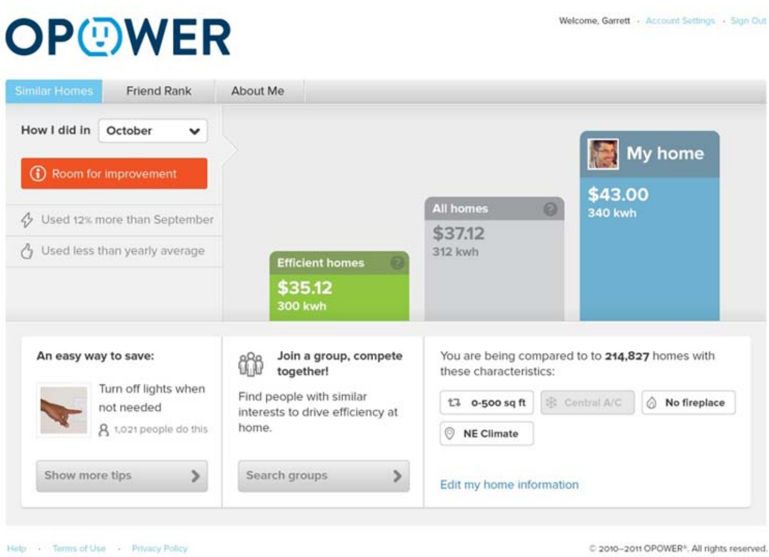
Big Data: Definition and Importance

What constitutes big data varies by time, application, and industry. Just as the number of transistors on an integrated circuit has grown exponentially per Moore's law, the amount of data captured and processed has increased exponentially with the digital revolution. Data can be considered big by virtue of its storage and transmission requirements or based on the computing power required to manipulate it. Big data is being collected by private enterprises including the telecommunications, scientific, and financial sectors, as well as by all levels of government.

From location data gathered from cell phones to digital images from airborne sensors to records of financial transactions, big datasets are growing and challenging our current data collection, data integration, and data analysis techniques. How data are formatted, stored, shared, and regulated affect intellectual property, scientific discovery, military effectiveness, and individual privacy.

Big data example:

OPOWER creates household-tailored data to inform customer of their energy use and how it compares to others in their neighborhood. OPOWER partners with utility companies to access vast amounts of energy usage data to create this information.



Opower, Facebook, the Natural Resources Defense Council (NRDC), and 16 U.S. utility companies create a social energy app to allow consumers to compare their energy usage with the average energy usage in their neighborhood. *"The new app is based on behavioral science…which has found that word-of-mouth is a good way to get people to save energy."*

Source: Clean Technica, http://s.tt/18LPb.

The sheer volume of big data being generated has consequences for data management and individual privacy. Where statute governs data sharing and privacy of data sources, in industries such as healthcare and financial services, big data can place a tremendous burden on data management and security resources. Even where data are not regulated, data security plays a critical role in big data, as any compromise to the data could result in negative consequences, including privacy or monetary losses to individuals or institutions and damages to the public's trust and confidence. Data stewardship—and the resulting standards, protocols, and legal requirements—will be important to ensuring that critical data are stored and shared in a secure manner. As datasets that need to be shared grow larger, distributed storage (so-called cloud computing) is becoming more the norm; however, newer approaches are likely to emerge.

A fundamental problem is that there currently are no realistic ways in these massive data systems to protect individual privacy and thus, serious legal mechanisms need to be put in place note only to control misuses of the data, but also to restrict access.

Big Data and Applied Social Science

Every global challenge faced today can be thought of in relation to data. Evidenced-based decision-making, understanding human behavior, and increasing amounts of data are coming together to create the next wave in social scientific research methods and analysis capabilities. Where typically social scientists attempted to collect data driven by some underlying research question, the result of which is known as a designed dataset (typical examples include data collected from surveys by statistical agencies), the pervasiveness of big data may create new opportunities for research previously thought to be impossible due to limitations in data collection.

The new paradigm is to collect all types of data without knowing immediately how they might be used or to save data collected for other purposes (such as cell phone data) that have the potential to be informative for other types of research. These types of data are sometimes referred to as organic data in contrast to traditional designed data of surveys and administrative data. See Table B-1 for examples of big data and what might be done with these independently generated data. The possibilities of accessing and manipulating data are becoming easier with the move to the cloud environment.

Table B-1. Data Are Growing Exponentially: Some Examples and Potential Uses

Organic data[a]	What might you do with these independently generated data?
<u>Location data</u> • Cell phone ”externals” • EZ pass transponders • Surveillance cameras <u>Political preferences</u> • Voter registration records • Voting in primaries • Political party contributions <u>Commercial information</u> • Credit card transactions • Real estate sales • Online searches • Radio-frequency identification <u>Health information</u> • Electronic medical records • Hospital admittances • Devices to monitor vital signs • Pharmacy sales <u>Critical infrastructure</u> • Transportation • Financial transactions • Power flow and transmission • Internet traffic <u>Other organic data</u> • Optical, infrared, and spectral imagery • Meteorological (e.g., temperature, pressure, wind, humidity, visibility, composition) • Mimetic, seismic, acoustic • Ionizing radiation, biological and chemical **Designed data** • Administrative data, (e.g., tax records) • Federal surveys • Census of population • Other data collected to answer specific policy questions	<u>Migration and location</u>[b] • Measure urban migration • Map population movements during natural disasters • Identify neighborhoods with inadequate social services • Map human behavior, such as dining-out habits, and correlate with health outcomes, such as diabetes <u>Critical infrastructure</u> • Develop rational infrastructure plans (e.g., traffic flow, zoning, public transit) • Examine the distribution and patterns of health events (e.g., disease surveillance and screening) <u>Energy related</u> • Practice monitoring, reporting, and verification for greenhouse gas emissions treaties • Detect hazards (e.g., leaks, plumes), emergency management • Establish energy efficiency standards for buildings and appliances • Use knowledge of behavior to encourage energy efficiency • Optimize grid operations <u>Methods and experiments</u> • Validate and calibrate proxies • Conduct policy experiments and simulations • Understand urban meteorology (e.g., leaks, plume dispersal) • Synthesize large seismic apertures for seismology and earthquake engineering

Sources: Adapted from Sallie Ann Keller, Steven E. Koonin, and Stephanie Shipp, “Big Data and City Living—What Can It Do for Us?” *Significance* (August 2012): 4–7. Inspired by Gary King, “Ensuring the Data-Rich Future of the Social Sciences,” *Science* 331 (6018), February 11, 2011: 719–21.

[a] Robert Groves, “Designed Data” and “Organic Data” (May 31, 2011), http://blogs.census.gov/directorsblog/2011/05/designed-data-and-organic-data.html.

[c] Alex Pentland, “Society’s Nervous System: Building Effective Government, Energy, and Public Health Systems,” *IEEE Computer Society* 45(1), 2012.

Categories of Stakeholders Interested in “Instrumenting Society”

Four categories of stakeholders have emerged as having a strong interest in accessing and analyzing big data. Their data, tools (analytical and theoretical), and practices vary widely and in some cases, may conflict. These four groups are not comprehensive in that they do not include the broader public as private citizens or other sectors more broadly, such as universities. However these four categories are likely to set the initial trends and encounter initial issues in big data management and utilization. Each is briefly described in the following subsections.

Private Sector

The private sector maintains multiple types of transactional and other data. Their interest is in monetizing the data they hold. Data kept by the retail sector may be used for targeted marketing and to encourage repeat customers; insurers use data to set rates. These data, correlated with other relevant data, potentially help to form a detailed picture of a time and place of critical financial interest to the enterprise.

A major issue for private sector stakeholders is that potentially valuable data are often kept only for specific periods of time for operational purposes, and then discarded. Some data are discarded because the storage requirements are prohibitive. For example, video data from a single surgical procedure may be retained for only a short time after the surgery. In other cases, data—surveillance footage, for example—may be discarded once they are deemed out of date. A second issue is that commercial enterprises many not want to share data if doing so aids competitors. They may, however, be less concerned with user privacy if they can sell or license user data.

Government and Other Social Service Providers

The public sector and other social service providers collect vast amount of survey, administrative, scientific, and other types of data, such as flight departure and arrival data. The scope and scale of government data being collected could possibly provide a more complete picture of a particular area's infrastructure, the health of a certain population, or the efficiency of provided services. Access and analysis of big data may also lead to better operations at the city level, including efficiency, resilience, equity, sustainability, quality of life.

However, restrictions can limit the ability to access data, aggregate with other datasets, and retain data after a set period of time. Both accessing and using big data and addressing the restrictions will require sophisticated computation, modeling, and simulation skills to create systems to inform decision-making.

Security Community

The security community (local law enforcement, Department of Homeland Security and FBI, and the Intelligence Community) has vast amounts of data that include electronic signals, satellite imagery, video streams, and sensor data, in addition to more traditional administrative data. Improvements in these data, such as advances in facial-recognition software, compression algorithms, and storage are likely to change investigative work. Other data streams, such as from utilities, Internet service providers, and telecommunications companies, are increasingly likely to be sought by law-enforcement personnel.

As was discussed in the body of this report, there are many additional issues for the security communities that need to be addressed. They are accumulating vast amounts of data yet do not have sufficient tools and capabilities to fully use the data. They have not resolved when and how such data can be shared among security agencies at all levels (Federal, State, and local). Other restrictions, such as the short length of time that the Intelligence Community is allowed to keep such data, also present challenges.

Social Scientists

For social scientists, access to big data fusing socioeconomic and physical data, offers a view of the social and physical environment in unprecedented detail. Such data could be used to understand why individuals, groups, and subpopulations act in certain ways and may further be applied to improving traffic flows, minimizing energy use, mitigating against certain crime patterns, among other benefits. Moreover, big data could help social scientists understand the effects of government policies on people's behavior. The ability to aggregate data from a variety of sectors could reveal hidden patterns of behavior, highlight economic trends, or reveal inefficiencies in infrastructure, among other uses.

These four stakeholder groups face similar challenges. If the benefits of big data are to be offset by the cost of their processing, data integration and analysis methods must evolve. Big data necessitates the development of new analytical methods to detect pattern disruptions and matches and build predictive models using extremely large datasets. Big data call for new statistical techniques that take into account large numbers of observations collected and structured in nontraditional ways. Novel algorithms that visualize data and communicate patterns in unique ways may also be required by these stakeholder groups. Important aspects of big datasets are that they may let you identify outliers more confidently; stratify more finely on correlative variables; assess and calibrate proxies more confidently; and calculate estimates for real time, very small populations/infrequent events, geographic detail.

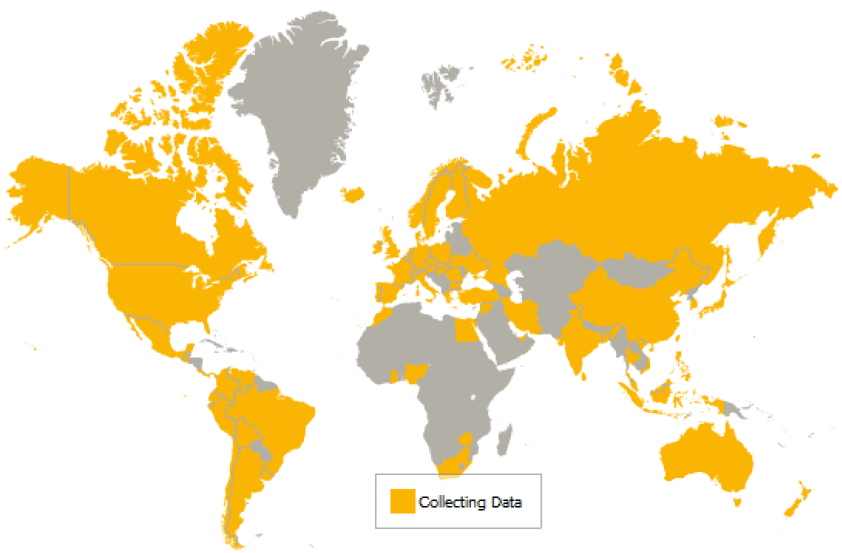
A key component of this effort is that the role of statistics will become even more central as traditional statistical methods are replaced by increasingly complex computational approaches. Engagement by quantitative methodologists will be necessary to provide infrastructural and computational guidance that enables the use of big data to solve societal challenges.

Big data example:

MIT Billion Prices Project (BPP)

This initiative uses prices collected from hundreds of online retailers around the world on a daily basis to conduct economic research. The figure below describes their most recent research leveraging high-frequency price data, as well as the U.S. daily inflation index.

BPP Geographic Coverage



Source: The Billion Prices Project @ MIT, http://bpp.mit.edu/.

This research uses high-frequency item-level data to study pricing topics in Macroeconomics to address:

Pricing Behavior: What drives price stickiness around the world? How much can be explained by current inflation, and inflation histories? How much by competition and industries' structure? Are prices synchronized?

Daily Inflation and Asset Prices: How do official statistics compare with daily inflation indexes across countries and sectors collected in this study? What are the links between daily inflation, asset prices, and inflation expectations?

Pass-Through: How much do prices adjust internally when the exchange rate or the international price of commodities change?

Green Markups: What premium is paid in stores for "green" or "organic" products?

Big Data Applied to Selected Social Science Topics

The study of smart cities, crime and social problems,[1] and civic engagement are three examples where the application of big data, along with the use of new tools and methods, could better inform policy.

Smart Cities

What if you could know everything about a city (state of its infrastructure, inhabitants, and environment) with high temporal and spatial resolutions? This would be achieved by fusing socioeconomic and physical data streams. What are the data streams that exist or could be generated? What data rates do they imply? How would they be acquired and analyzed? And what could you do with them?

These are questions largely unexplored by the social scientists studying cities, as their traditions are either high-purity data collected by surveys numbering in the thousands, or statistical data flows that sample widely, but offer little opportunity for segregation. Both these methodologies are also non-real-time and they are expensive. A complementary strategy is to use new technologies to collect and analyze far greater volumes of physical and socioeconomic data. The quality will be less than is traditional, but corrections can be applied. Such approaches are familiar in astronomy, remote sensing, and biology. Their utility in the social sciences is just beginning to be explored.

A smart city is one that provides information and reacts to information. Smart cities can be characterized at three levels. The first passively collects data. The second informs stakeholders to make tactical decisions based on these data, and the third makes strategic decisions based on the data automatically.

Globally, a key input to the design of smart cities could be the United Nations' Millennium Development Goals which include the provision of universal education, the end of poverty and hunger, gender equality, child health, maternal health, measures to combat HIV/AIDS, environmental sustainability to optimize scarce resources, the development of global partnerships for development, and freedom of choice (which implies equal access).

For a smart city to be successful there must be incentives for the citizens to allow access to data, such as better energy management and improved traffic patterns. At the same time, city managers will have to decide what approaches to take in utilizing the

---

[1] One participant objected strenuously to grouping the vast array of social trends/challenges under "crime," pointing out that this missed interesting applications of big data to social policy. The participant felt that the portrayal of the discussion much better reflected the broader perspective that big data may be used to ensure equity and distributive justice, rather than focusing on crime and social problems.

data: Will a smart city change people's behavior, or will its systems adapt to people's behavior? Developing a smart city requires improved data streams. But these data raise a host of issues, such as:

- Who owns the data and how much information is needed?
- What are the consequences for misuse of data?
- What are the institutional barriers to integrating these streams?
- What are the privacy concerns? For example, data could be linked through facial recognition and geo-referencing, but doing so raises privacy concerns.

While there is a body of knowledge and practice in urban studies, new technologies offer opportunities to understand what goes on in cities in greater detail and scope. Optimal use of big data can lead to understanding the city's operation, and hence new possibilities for optimization.

Crime and Social Problems

Big data may be used to ensure equity and distributive justice, rather than focusing on crime and social problems. The question is "What can big data do to improve the way urban communities function?"

There may of course be several perspectives to answer this question. For example, big data could be used as an early-warning indicator of neighborhood change. Such indicators may be in the form of recent housing-stock sales data or crime data. Big data could address health issues. Some studies are looking at correlations between transportation routes and asthma, and data could be collected and studied to determine what works and what does not as alternatives to managed care. Big data could inform law-enforcement policies by providing better indicators of deterrence, more accurate data on crime rates, and as a measure of efficiency (does the money spent on a particular law-enforcement action justify the cost?).

At present, legal issues regarding how data are shared are in flux, and the capacity for surveillance and monitoring is only going to grow. Regulatory guidance is necessary to ensure that data are not used for nefarious purposes, such as sophisticated redlining of neighborhoods. As more big data become available, easier and broader access to them will be demanded. Balancing privacy rights against security needs will likely be played out in the courts for some time.

Civic and Public Participation

Big data could address three main questions related to civic and public participation:

1. How do you measure engagement using big data?

2. How do you use big data measurements to engage stakeholders?
3. How do you involve stakeholders in using big data as part of civic engagement?

Some practical applications amplify the question. For example, consider lists of potential jurors and the varieties of ways these are managed by jurisdiction. Could big data be used to characterize participation or to better manage the list? Similarly, could big data characterize the not-in-my-backyard phenomenon? Civic engagement is difficult to measure because of the different forms it takes, such as voting, volunteering, or protesting. And it is not always clear what is being observed. Is a person photographed at a protest a participant or someone passing through the area? If you know everything about an individual you might have a profile of engagement, but achieving that knowledge raises a host of privacy questions.

Some current projects are attempting to study civic engagement by gathering data from multiple sources. There are several examples.

- The New York University Center for Urban Science and Progress (CUSP) will be a test bed and a living laboratory analyzing data from sources measuring the physical condition of the city, its operation, and its people.
- Intel is partnering with Imperial College and University College London to set up an urban laboratory, similar to CUSP. The laboratory's agenda will be citizen-led and implemented by twelve researchers that include social scientists, computer analysts and others.

Challenges with Using Big Data

Retrieving and analyzing increasingly massive amounts of data to inform decision-making will require new architecture and computational statistical theories and applications. Detecting patterns in data and translating them to information outputs that can be used quickly and without ambiguity will become increasingly difficult as the size of datasets collected continues to grow.

Industry will require new types of jobs including data service providers (to standardize, curate, and make data available), data analysts (statisticians, mathematicians, computer scientists, computational social scientists, and others to develop the computational methods), and data analytic managers (to ask pertinent research questions and extract meaning from the data). Meeting the demand for workers skilled in the research and use of big data requires improvements in education at all levels, including developing workers who are deeply trained in informatics across many disciplines. This will require the need for public-private partnering at local, state, and country levels to incorporate changes into K-12, community college, university, and workforce training.

Privacy concerns arise on several fronts. Data privacy has financial implications to individuals and institutions if the data are lost, stolen, or inappropriately shared. Institutions can lose valuable research results or face legal claims for damages for improperly secured data. Massive data retention such as that by government organizations creates opportunities to recognize patterns and extract meaning from data, yet security requirements must be balanced against privacy concerns. However, more discussion is needed. This discussion must focus on data ownership about individuals collected from all the various social media, transactional, and surveillance sources.

Advances in the development of data collection techniques, tools, and methods are currently hampered by a lack of access to large datasets and by lack of funding. The National Science Foundation has shown interest in quantitative behavioral and social science, which can dramatically augment traditional survey instrumentation. Instrumentation and methods should ideally be coupled with theory. Again, privacy and confidentiality issues are important considerations in this regard.

Handling massive amounts of data will require standards and protocols for their storage and archiving, intellectual-property rules and data-use agreements, and computational statistical advancement. Replicated results and shared data are necessary for advancing science; these should be encouraged by the academic community and practiced by research institutions. Common standards for scientific data support their preservation and reuse and make interdisciplinary work easier to accomplish.

Most often, quantitative social scientists traditionally used or designed datasets for specific purposes, but the availability of big data and the opportunity to instrument entire subpopulations to create new sources of data may represent a shift in methodological and analytical techniques. Statistical sampling, where a part is measured to represent the whole, arose as an alternative to surveying all members of a group, which was often prohibitively expensive. As big data become less expensive to procure and more prevalent in their use, traditional statistical sampling techniques may no longer apply.

If big data become available to researchers in such unprecedented quantities, the question becomes when is there enough density of data for “natural experiments”—a means to determine how behavior change may be possible without resorting to the expensive clinical trials system. Researchers need to develop techniques that obscure individual identities. And simply having access to big data does not equate to value. Thinking systematically about how to use the data and about what structures should be in place for analyzing them is necessary. After the initial determination of patterns (descriptive analyses), researchers must generate hypothesis or research questions and then test and answer them.

Big data from urban areas may reflect social networks of employment, education, civic engagement, and others. Determining the characteristics of these networks, membership, behavior, function, could lead to important social science applications. For example it could reveal how a disease spreads and in turn lead to policies to mitigate its spread. However, the tools to collect and analyze these data remain to be developed, and the privacy issues remain to be resolved.

Data Acquisition

If big data are collected by disparate groups, incentives are needed to promote data sharing. Gaining access to industry data, which are often closely held for competitive advantage, may require offering a service to the owner. Some companies aggregate data from public records. In cases where they are not willing to share data, the data may be acquired from the original sources, albeit at a cost. Where the data are sole source, there may not be an alternative if the company is unwilling to share.

Other obstacles to sharing may remain. Terms of use on individual websites, along with laws and regulations that vary by jurisdiction, may prohibit or sharply limit data sharing. Some data cannot be shared for security reasons. There are a number of models for data sharing, but scaling them up to the big data world is difficult.

At the same time, some industries may be open to sharing their big data because they are just starting to think about how to use what they have. Social scientists should look for ways to leverage alliances between large companies with big data, try to gain inroads into industry data by highlighting their trustworthiness, and cite the advantages of data sharing to the organization with the data. The potential risk is that independence and objectivity may be undermined if researchers are dependent on corporations for continued access to data.[2]

Data Curation

Curation refers to how data are formatted and stored. Big data contradict the old notion of data as static records, frozen in time—big data are dynamic, with every cell in a database having a history. But in the United States, there are no standards for data curation; standards are only emerging in other countries. Complicating this, how an organization uses data is tied primarily to the needs and obligations of the organization itself; at a deeper level, the format and handling of the data may be the responsibility of the programmers doing the coding.

---

[2] In reviewing this workshop summary, one participant commented that this is already a problem with the companies that own data on mortgage loan performance. "If you publish results they don't like, they won't sell you the data next time."

Incentives for curation may exist if curation can be shown to improve data security, but there are often institutional tensions on how to manage data. Data curation is not an easy issue to solve, but it should be at the forefront of any analytical enterprise.

Data Integration

Data integration is as important as data analysis, but it is only being done incrementally. Like data sharing, data integration is complicated by restrictions on datasets and incomplete standards. Because data come in multiple forms besides ASCII, for which metadata standards exist, new standards that accommodate digital, among other data, need to be developed if such data are to be successfully integrated.

Data Analysis

The analogy to finding the needle in the haystack has relevance here; only in the case of big data, the shape of the haystack may be as important as the needle. The analysis of big data can reveal previously hidden features, such as new markets, and they allow descriptions of normal patterns (an analysis of the haystack). Beyond these patterns, the residuals—outliers, secondary effects—may prove more interesting (finding the needle in the haystack). Advanced analytical techniques could help reveal their meaning. Similar challenges arise from the haystacks of information on social network, time-use, economic, and other activities that can be mined to benefit science, business, and society, as well as to identify and calibrate proxies when data are not available for specific groups or variables.

Some caveats to the analysis of big data must be considered. If subjects know that data are being collected about them, does that influence their actions? If data collected are used to evaluate performance, does that affect how employees do their jobs? In other words, what are the observation effects? Last, the digital divide means there is a demographic bias in the data sampled, which means that big datasets are not representative.

Ethical and Legal Considerations

Ethical and legal considerations for users of big data include ownership of data, ethical and legal responsibilities of researchers, and how to deal with personally identifiable information. For example, suppose routine review of digital imagery reveals a health problem. Whether this should be reported must be determined before the review even begins. Analysis of imagery from urban centers could reveal criminal activity. Are the researchers obligated to report it? What contingencies on reporting might a firm specify before allowing its facilities to be monitored? Anticipating the likely uses of gathered data and the patterns they may reveal is nontrivial.

Replicability

Recreating results obtained from analyses of big data is necessary to validate social science experiments. However, replicating results of research that uses big data is not straightforward because the underlying data are not static and because the tools used to validate the data may no longer be useful. Terms of use and other restrictions to data access can also make replicability difficult. Some universities have formed collaborations to access big data, such as those that enable them to use census data. In some cases, a secure environment was required, and once the experiment concluded, the underlying data had to be destroyed.

A second facet of replication is the idea that with enough data, a digital representation of what is going on in society is possible. If these data can be synthesized, there can be digital experiments, in theory. For example, such data might better reveal the characteristics of a market, for example, the effects of changing prices and demographic trends on purchasing.

In summary, one of the promises of big data is an unprecedented level of detailed information, which could lead to better analysis and a different understanding of individual or group behavior as well as government or institutional function. Achieving this promise requires resolving issues related to acquiring, curating, integrating, and analyzing data; recognizing the ethical and legal considerations that apply; and ensuring that results from big data can be validated through replication.

Summary and Recommendations for Next Steps

Four topical themes emerged during the workshop:

- *Big data may shift traditional quantitative social science methodologies from small-scale experiments and surveys to actual (not modeled) population-based observations*. Such a shift will likely have an effect on current statistical techniques including a reliance on a self-selected census of units rather than a sample of units with precise coverage.
- *At least four categories of interested stakeholders are interested in instrumenting society:* the private sector, government and related organizations, the security community, and social scientists. The role of citizens was not discussed directly.
- *Big data can be applied to issues facing society in myriad ways yet face similar challenges.* Three examples of the application of big data to addressing social science issues are (1) creating smart cities, (2) addressing crime and social problems, and (3) encouraging civic engagement.

- *Challenges with using big data in social science research* include privacy concerns in addition to methods of data acquisition, curation, integration, and data analysis, ethical and legal considerations, and data replicability.

A number of steps are recommended to further the use of big data to address societal issues:

- Establish standards for usage – including for ensuring individuals' privacy – from the start rather than attempting to do after widespread use of big data permeates
- Conduct research to understand optimal approaches for accessing and integrating data, especially data from multiple sources and locations.
- Identify the social and behavioral science questions that would benefit from using big data.
- Continue to fund public-private initiatives to develop computational statistical approaches to detect patterns and produce outputs that support real-time decision-making.
- Create a strategy for education and workforce training to integrate, standardize, and extract wisdom from all types of data.
- Develop a common, open-source, collaborative infrastructure to facilitate data analysis and sharing; such an infrastructure should be interdisciplinary across scholarly fields.
- Create a community of advisors and practitioners who regularly meet to discuss policy perspectives on the uses and challenges associated with using big data.

References

Abelson, Hal, and Lawrence Lessig. "Digital Identity in Cyberspace." White Paper submitted for 6.805/Law of Cyberspace: Social Protocols. December 10, 1998. http://groups.csail.mit.edu/mac/classes/6.805/student-papers/fall98-papers/identity/linked-white-paper.html.

Albro, Robert. "Minerva and Critical Public Engagement." The Social Science Research Council, The Minerva Controversy essays, 2008. http://essays.ssrc.org/minerva/2008/11/14/albro/

Albro, Robert, James Peacock, Carolyn Fluehr-Lobban, Kerry Fosher, Laura McNamara, George Marcus, David Price, Laurie Rush, Jean Jackson, Monica Schoch-Spana, and Setha Low. *Final Report on The Army's Human Terrain System Proof of Concept Program*. AAA Commission on the Engagement of Anthropology with the US Security and Intelligence Communities (CEAUSSIC), October 14, 2009. http://www.aaanet.org/cmtes/commissions/ceaussic/upload/ceaussic_hts_final_report.pdf.

Al-Hader, Mahmoud, and Ahmad Rodzi. "The Smart City Infrastructure Development & Monitoring." *Theoretical and Empirical Researches in Urban Management*, 2 (11, 2009): 87−94.

Allwinkle, Sam, and Peter Cruickshank. "Creating Smarter Cities: An Overview." *Journal of Urban Technology* 18 (2, 2011): 1−16.

Asher, Thomas. "Making Sense of Minerva Controversy and the NSCC." The Social Science Research Council, The Minerva Controversy essays. October 9, 2008. http://essays.ssrc.org/minerva/2008/10/09/asher/.

Barton, Dominic, and David Court. "Making Advanced Analytics Work for You." *Harvard Business Review*, October 2012. Reprint R1210E: 7.

Bergemann, Richard J., and Nancy P. Licato. *Summary Report of the Defense Science Study Group 1985–1988*. IDA Paper P-2310, September 1989. http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA217607.

Broad, William J. "U.S. Will Deploy Its Spy Satellites on Nature Mission." *The New York Times*, November 27, 1995.

Brumfiel, Geoff. Shared Intelligence. *Nature* 477 (2011): 388−389.

Cloud, John G. and Clarke, Keith C. "Through a Shutter Darkly: The Tangled Relationships between Civilian, Military and Intelligence Remote Sensing in the Early U.S. Space Program." In *Secrecy and Knowledge Production* edited by J. Reppy, 132. Peace Studies Program Occasional Paper #23, Cornell University, October 1999.

Committee on the Scientific Value of Arctic Sea Ice Imagery Derived Products; Committee on Climate, Energy, and National Security; National Research Council. *Scientific Value of Sea Ice Imagery Derived Products*. Washington, DC: National Academies Press, 2009: 48.

Davenport, Thomas H., and D. J. Patil. "Data Scientist: The Sexiest Job of the 21st Century." *Harvard Business Review*, October 2012, reprint R1210D: 8

Deibert, Ronald J. "From Deep Black to Green? Demystifying the Military Monitoring of the Environment." In *Woodrow Wilson Center Environmental Change and Security Project Report*, Washington, DC, Spring 1996.

Engerman, David. "Knowing the Enemy," The Social Science Research Council. The Minerva Controversy essays, 2009. http://essays.ssrc.org/minerva/2009/01/23/engerman/.

Groves, Robert. " 'Designed Data' and 'Organic Data' " United States Census Bureau Director's Blog. May 31, 2011. http://blogs.census.gov/directorsblog/2011/05/designed-data-and-organic-data.html.

Harrison, Colin, and Ian Abbott Donnelly. "A Theory of Smart Cities." In *The Proceedings of the 55th Annual Meeting of the International Society for Social Science*. United Kingdom: University of Hull Business School, July 17–22, 2011.

Hodgson, Bob. "Technopolis: Challenges and issues: A Tale of (at Least) Three Cities." In David V. Gibson, George Kozmetsky, and Raymond W. Smilor (eds.), *The Technopolis Phenomenon: Smart Cities, Fast Systems, Global Networks*. Lanham, MD: Rowman & Littlefield Publishers, Inc., 1992.

Jung, John G. "Smart Communities: Digitally-Inclined and Content-Rich." *New Telecom Quarterly* (1Q, 1998): 19−26.

Keller, Sallie Ann, Steven E. Koonin, and Stephanie Shipp, "Big Data and City Living—What Can It Do for Us?" *Significance* (August 2012): 4–7.

King, Gary. "Ensuring the Data-Rich Future of the Social Sciences," *Science* 331 (6018, February 11, 2011): 719–21.

Krebs, Robert R. "Minerva: Unclipping the Owl's Wings." The Social Science Research Council, The Minerva Controversy essays, 2008. http://essays.ssrc.org/minerva/2008/11/19/krebs/.

Mahizhnan, Arun. "Smart Cities—The Singapore Case." *Designed Cities* 16 (1, 1999), 13−18.

Mahnken, Tom. "Building Bridges and Communities." The Social Science Research Council, The Minerva Controversy essays, 2008. http://essays.ssrc.org/minerva/2008/12/30/mahnken/.

McAfee, Andrew, and Erik Brynjolfsson. "Big Data: The Management Revolution." *Harvard Review* (October 2012, reprint R1210c): 9.

McFate, Montgomery, and Steve Fondacaro. "Reflections on the Human Terrain System during the First 4 Years," National Defense University, *Prism* 4 (2, September

2011). http://www.ndu.edu/press/lib/images/prism2-4/Prism_63-82_McFate-Fondacaro.pdf.

Nam, Taewoo, and Theresa A. Pardo. "Conceptualizing Smart City with Dimensions of Technology, People, and Institutions." In *The Proceedings of the 12th Annual International Conference on Digital Government Research*, June 12–15, 2011, College Park, MD.

Neville, Warwick. "Managing the Smart City-State: Singapore Approaches the 21st Century." *New Zealand Geographer* 55 (1, 1999): 35−45.

National Research Council. Committee on National Security Implications of Climate Change for U.S. Naval Forces. *National Security Implications of Climate Change for U.S. Naval Forces*. 2011. http://www.nap.edu/catalog.php?record_id=12914.

National Science Foundation. *Changing the Conduct of Science in an Information Age*. Arlington, VA, 2011. http://www.nsf.gov/pubs/2011/oise11003/.

National Task Force on High Energy Density Physics. *Frontiers for Discovery in High Energy Density Physics*. Prepared for Office of Science and Technology Policy National Science and Technology Council Interagency Working Group on the Physics of the Universe, 2004. http://www.whitehouse.gov/sites/default/files/hedpreport.pdf.

Office of the Secretary of Defense for Acquisitions, Technology, and Logistics (ASDR&E), *Report of the Defense Science Board Task Force on Basic Research*, Washington, D.C.: ASDR&E, January 2012.

Pace, Scott, Kevin M. O'Connell, and Beth E. Lachman. *Using Intelligence Data for Environmental Needs—Balancing National Interests*. Santa Monica, CA: RAND National Security Research Division, 1997.

Pentland, Alex. "Society's Nervous System: Building Effective Government, Energy, and Public Health Systems," *IEEE Computer Society* 45 (1, 2012).

Petras, James. "Guest Editorial, Procuring Academics for Empire: The Pentagon Minerva Research Initiative." *Dialect Anthropol* 33 (2009): 1−4.

Ratti, Carlo, and Anthony Townsend. "The Social Nexus." *Scientific American* 305 (2011): 42−48.

Richelson, Jeffrey T. "Scientists in Black." *Scientific American* (February 1998): 48−55.

Simpson, Christopher. *Science of Coercion, Communication Research & Psychological Warfare, 1945–1960*. New York: Oxford University Press, 1994.

Snyder, Diane, and Michael L. Brown. "The Intelligence Community and the Environment." In *Intelligence Reform in the Post-Cold War Era*," WWS 401a, 1997.

Weinberg, Steven. "The Crisis of Big Science," *New York Times Review* (May 10, 2012).

Abbreviations

ATP	Advanced Technology Program
DSB	Defense Science Board
CUSP	Center for Urban Science and Progress
DARPA	Defense Advanced Research Projects Agency
DOD	Department of Defense
DSSG	Defense Science Study Group
IARPA	Intelligence Advanced Research Projects Activity
IC	Intelligence Community
IDA	Institute for Defense Analyses
MEDEA	Measurements of Earth Data for Environmental Analysis
NIST	National Institute of Standards and Technology
SAB	Scientific Advisory Board

REPORT DOCUMENTATION PAGE			*Form Approved OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
XX-05-2013	Final	Mar 2012 - May 2013

4. TITLE AND SUBTITLE

Pathways to Cooperation between the Intelligence Community and the Social and Behavioral Science Communities

5a. CONTRACT NUMBER

DASW01-04-C-0003

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Koonin, Steven E.
Keller, Sallie A.
Shipp, Stephanie S.
Allen,Ted W.
Walejko, Gina K.

5d. PROJECT NUMBER

5e. TASK NUMBER

ET-20-3393

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Institute for Defense Analyses
4850 Mark Center Drive
Alexandria, VA 22311

8. PERFORMING ORGANIZATION REPORT NUMBER

IDA Paper P-5000

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

National Intelligence Officer for Science & Technology
National Intelligence Council
Office of the Director of National Intelligence
OHB 2E42
Washington, DC 20511

10. SPONSOR/MONITOR'S ACRONYM(S)

ODNI

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited. (May 14, 2013)

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The U.S. Intelligence Community (IC) would like to collaborate with civilian social and behavioral scientists in threat detection, assessment, deterrence, and minimization. Social and behavioral scientists can provide crucial expertise and support in analyzing, interpreting, and predicting potential threat behaviors and organizational development and in assessing the social vulnerabilities of U.S. infrastructure. This infrastructure is increasingly reliant on "big data" derived from digital technology that can be used to assess threats. Collaborations with social and behavioral scientists will allow the IC to build analysis capacity in new directions and with new communities. This will allow the IC to build future capability that they can use to address unknown problems and integrate information in new, yet-to-be-determined ways.

15. SUBJECT TERMS

Big Data, Minerva Research Initiative, Measurements of Earth Data for Environmental Analysis (MEDEA), Intelligence Advanced Research Projects Activity (IARPA), Defense Sciences Study Group (DSSG), collaborations, social and behavioral science communities, smart cities

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Gershwin, Lawrence K.
Unclassified	Unclassified	Unclassified	Same as Report	62	19b. TELEPHONE NUMBER *(Include area code)* 703-482-6811